



**Europäisches  
Patentamt**

**European  
Patent Office**

**Office européen  
des brevets**

30. 07. 2004

REC'D 12 JAN 2005

WIPO

PCT

**Bescheinigung**

**Certificate**

**Attestation**

Die angehefteten Unterla-  
gen stimmen mit der  
ursprünglich eingereichten  
Fassung der auf dem näch-  
sten Blatt bezeichneten  
europäischen Patentanmel-  
dung überein.

The attached documents  
are exact copies of the  
European patent application  
described on the following  
page, as originally filed.

Les documents fixés à  
cette attestation sont  
conformes à la version  
initialement déposée de  
la demande de brevet  
européen spécifiée à la  
page suivante.

**Patentanmeldung Nr. Patent application No. Demande de brevet n°**

03450184.1

**PRIORITY  
DOCUMENT**

SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;  
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets  
p.o.

R C van Dijk

**BEST AVAILABLE COPY**



Anmeldung Nr:  
Application no.: 03450184.1  
Demande no:

Anmeldetag:  
Date of filing: 11.08.03  
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Greenovation Biotech GmbH  
Boetzingen Strasse 29 B  
79111 Fribourg  
ALLEMAGNE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:  
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.  
If no title is shown please refer to the description.  
Si aucun titre n'est indiqué se référer à la description.)

Moss expressing promoting regions

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s)  
revendiquée(s)

Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/  
Classification internationale des brevets:

G12N15/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of  
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL  
PT RO SE SI SK TR LI

The invention relates to isolated nucleic acid molecules promoting expression of polypeptides in genetically modified eukaryotic host cells.

The expression of proteinaceous substances (proteins, peptides, polypeptides, fragments thereof, as well as posttranslationally modified forms of these molecules are hereinafter referred to as "polypeptides" (synonymously used together with "protein", e.g. in the example part) in genetically modified cells is a major source for providing preparations of such often rare and valuable substances. For expressing such polypeptides in genetically modified host cells, the presence of a DNA region is necessary which positively controls ("activates", "promotes") this expression. Promoters are important examples for such regions allowing RNA polymerases to bind to the DNA for initiating transcription into mRNA (Watson et al., "Recombinant DNA" (1992), Chapter I.1 and 2).

Mosses have gained increasing attention as useful objects for research for plant physiology and development, since their simple nature (mosses are situated at the base of higher-plant-evolution) provides insights into the complex biology of higher plants. The simple morphology of mosses and the advantageous culturing possibilities has made them popular model organisms for studies of plant physiology and developmental biology: Moss species may be cultured without difficulty under controlled conditions, using in vitro techniques including axenic culture, not only in petri dishes, but also in liquid culture e.g. in bioreactors. The haploid gametophyte can be grown photoautotrophically in sterile culture and easily observed at the cellular level.

Another major advantage of mosses is their transformation capacity: Despite numerous studies, the ratio of targeted integration events in plants hardly reaches  $10^{-4}$ , which prevents the general use of gene targeting approaches for plant functional genomics. In contrast to all other plants having been tested so far, integration of homologous DNA sequences in the genome of mosses (especially the established moss model organisms such as *Physcomitrella patens* (for a review of its molecular genetics:

Reski, 1999)) occurs predominantly at targeted locations by homologous recombination. Transformation of mosses is usually and easily performed via PEG-mediated uptake of plasmid DNA by protoplasts, DNA transfer by microprojectile bombardment, electroporation and microinjection (Cove et al., 1997). Depending on the design of the transforming construct predominantly random or targeted integration occurs.

Despite the use of mosses as scientific tools for plant physiology research, the use of mosses for producing recombinant heterologous polypeptides in moss cells has been rather limited so far, although efficient production methods have become available (e.g. culturing protonema moss tissue as described in EP 1 206 561 A).

A major limitation of transformation technologies in eukaryotic host cells, especially in animal cells or cells of higher plants, has always been the lack of an efficient promoter for high constitutive expression of foreign genes in such transgenic host cells. The cauliflower mosaic virus (CaMV) 35S promoter has been widely used for this purpose in a number of plant transformation systems, however, the CaMV 35S promoter has shown low activity in some plant species (specially monocots, such as rice (McElroy et al., 1991,)). For monocot transformation the rice actin 1 5' region has been used for heterologous expression of proteins (McElroy et al., 1991,). Nevertheless, the continuing need to provide novel expression promoting means for the expression of recombinant (foreign) polypeptides in genetically modified eukaryotic host cells still exists.

For mosses, especially for *Physcomitrella patens*, up to now, no homologous (in this case homologous is defined as: moss derived) suitable nucleus derived expression promoters or other nucleus derived expression promoting sequences have been published so far (Holtdorf et al., 2002). Researchers have therefore used heterologous (in this case heterologous is defined as: not moss derived) promoters for the expression of selection marker genes and other genes of interest. However, only a few of such promoters have been reported to function reliably in certain mosses (e.g. the CaMV 35S-promoter; summarised in Holtdorf et al.,



2002; CaMV 35S-promoter does not work in certain other species (Zeidler et al., 1999)). Therefore, other means for genetically manipulating mosses have been developed in the art, e.g. gene-trap and enhancer trap systems (Hiwatashi et al., 2001; however, also using CaMV 35S promoter).

Whereas in the above mentioned research in mosses using homologous recombination the use of heterologous promoters is necessary (and therefore homologous promoters are not needed, moreover they are in most cases not useful), the need for a suitable moss derived expression promoting means for industrially using mosses for the production of recombinant polypeptides or for the overexpression of homologous polypeptides is present and yet unsolved. Such expression promoting means should allow a stable and constitutive expression under the applied culturing conditions and should preferably enable a comparable or even higher expression performance as the CaMV 35S promoter.

Therefore, the present invention provides an isolated nucleic acid molecule encoding a moss expression promoting region (MEPR). With the present invention moss derived expression regions are provided which allow a constitutive expression in genetically modified host cells, especially mosses, thereby addressing the needs for such tools raised in the prior art (Holtdorf et al., 2002; Schaefer et al., 2002).

An essential feature of the MEPRs according to the present invention is also that the expression promoting activity of the MEPRs is at least 30 %, preferably at least 50 %, of the expression promoting activity of a working heterologous promoter in the specific host cell (e.g. CaMV 35S for the expression of a recombinant polypeptide in *Physcomitrella patens*), because moss promoters which do not have such an expression promoting activity cannot be properly used for solving the objects of the present invention and are therefore not regarded as MEPRs.

According to the present invention MEPRs promoting expression which is not sphorophyte specific, are defined as constitutive MEPRs, preferably MEPRs promote expression in gametophyte derived cells, more preferably MEPRs promote expression in protonema cells.

According to the present invention constitutive expression is preferably defined as the expression of a protein resulting in detectable amounts of this protein under liquid culture conditions generally used for photoautotrophically grown mosses, e.g. flask cultures, bioreactor cultures (EP 1 206 561 A), conditions used for the transient expression system described beneath. Therefore, constitutive expression has to be given for the MERPs according to the present invention preferably without the need of specific culturing additives, preferably also without the need of added sugars, phytohormones or mixtures of such substances in the culture medium. The constitutive expression has to be performed in a steady mode; yet it can be transient.

The terms "moss" or "mosses" as used in the present specification encompasses all bryophytes (hepatics or liverworts, hornworts and mosses). Characteristic for mosses is their heteromorphic *Generationswechsel*, the alternation of two generations which are distinct from each other in terms of nuclear DNA amounts and morphology. The diploid sporophyte is photosynthetically active only in its youth and requires supply from the dominating, green, haploid gametophyte. The gametophyte exists in two morphologically distinct forms: the juvenile gametophyte, called protonema and the adult gametophyte, called gametophore. In contrast to the protonema, the adult gametophyte (gametophore) bears the sex organs.

In the context of the presented invention transient expression is defined as introduction of an episomal nucleic acid-based construct (e.g. MEPRs and gene of interest) as described below into a moss protoplast and causing or allowing transient expression from the vector that results preferably in turn to the secretion of extracellular protein into the medium. Protoplasts are derived from moss cells, preferably, from gametophytic cells, more preferably from protonema cells.

Although the MEPRs according to the present invention may be taken from any moss species, the MEPRs are preferably isolated from common model moss species. The MEPRs are therefore preferably isolated from *Physcomitrella*, *Funaria*, *Sphagnum*, *Ceratodon*,

Marchantia and Sphaerocarpos, especially of Physcomitrella patens, Funaria hygrometrica and Marchantia polymorpha.

Suitable MEPRs according to the present invention are selected from the Seq. ID Nos. 1 to 27 or expression promoting fragments thereof. An "expression promoting fragment" is a fragment of an MEPR which has an expression promoting activity of the MEPRs of at least 30 %, preferably at least 50 %, of the expression promoting activity of a working heterologous promoter in the specific host cell (e.g. CaMV 35S for the expression of a recombinant polypeptide in Physcomitrella patens).

The MEPRs according to the present invention may comprise specific regions, such as a promoter region ("promoter"), 5'untranslated regions ("5'-UTRs"), 5'-introns or 3'-UTRs. For some MEPRs, expression promoting fragments exist which only contain the 5'-intron. Usually the promoter is always active alone as an expression promoting fragment. Therefore, the MEPR according to the present invention preferably comprises a moss promoter and preferably a 5'-UTR region and/or a 5'-intron and/or a 3'-UTR .

Although it is often sufficient, if a certain constitutive expression is reached, it is in many cases preferred to achieve a high expression rate, especially for industrially producing recombinant polypeptides. Most of the MEPRs according to the present invention have proven to allow significantly higher expression rates for a given recombinant polypeptide than the CaMV 35S promoter, especially in homologous systems (e.g. a Physcomitrella MEPR for expression of a polypeptide in Physcomitrella). Therefore, preferred MEPRs according to the present invention have an expression promoting activity being at least equal to the expression promoting activity of cauliflower mosaic virus (CaMV) 35S promoter, especially, but not limited, in the moss species from which the MEPR was isolated. Even more preferred MEPRs have an expression promoting activity being at least 200 %, preferably being at least 500%, especially being at least 1000 %, of the expression promoting activity of cauliflower mosaic virus (CaMV) 35S promoter, especially, but not limited, in the moss species from which the MEPR was isolated.

The isolated nucleic acid molecules according to the present invention are preferably used to transform a specific host cell for producing a recombinant transgenic polypeptide, preferably, but not limited to, in an industrial scale. Therefore the nucleic acid molecule is provided as a suitable vector allowing transformation and expression of the transgene in the host cell. Among the possibility that an MEPR according to the present invention is used for replacing a natural promoter in mosses, thereby bringing the expression of a homologous moss polypeptide under the control of a MPER being located at a position in the genome of the moss, where it is normally not present in wild type strains, the prevalent industrial applicability of the present MEPRs is the control of expression of a heterologous ("foreign") gene in a production host cell, specifically a plant cell, especially a moss cell. Therefore, the nucleic acid molecule according to the present invention further comprises a coding region for a recombinant polypeptide product, said coding region being under the control of the MEPR.

It is also advantageous, if the isolated nucleic acid molecules according to the present invention further comprises a selection marker and/or further regions necessary for enabling the appropriate transformation method chosen (see e.g. Cove et al., 1997; Schaefer, 2002). For example, if targeted integration is preferred, the nucleic acid molecule according to the present invention should further comprise sequences which are homologous to genomic sequences of the species to be transformed. Thus, allowing targeted integration of the isolated nucleic acid molecule via homologous recombination into the genome of the species to be transformed.

Moreover, the isolated nucleic acid molecules according to the present invention can be used for screening and defining consensus sequences for expression promoting regions. Finding and screening for such consensus sequences (regions, boxes) which are important and/or essential for expression promoting activity is a valuable asset in recombinant DNA technology, especially with respect to industrial biotechnology using mosses.

According to another aspect, the present invention also relates to a process for the expression of a recombinant polypeptide product in an eukaryotic host cell comprising the following steps:

- providing a recombinant DNA cloning vehicle comprising an isolated nucleic acid molecule encoding an MEPR according to the present invention and optionally a coding region for said recombinant polypeptide product, said coding sequence being under the control of the MEPR of said nucleic acid molecule in said host,
- transforming said eukaryotic host cell which does not naturally harbour said coding sequence in a way that it is under the control of said MEPR,
- culturing the transformed eukaryotic host cell in a suitable culture medium,
- allowing expression of said recombinant polypeptide and
- isolating the expressed recombinant polypeptide.

As mentioned above, MEPRs according to the present invention in principle have the capability to achieve constitutive expression in various cell types, the eukaryotic host cell is preferably selected from plant cells, preferably moss cells, especially *Physcomitrella patens* cells.

A system which is specifically preferred for the present invention is the culturing in moss protonema cultures (protonema moss tissue). In doing so the method described in the EP 1 206 561 A and the preferred embodiments thereof are explicitly incorporated by reference herein and are immediately applicable to the present invention.

The constitutive expression of the polypeptide with the means according to the present invention is possible without the need for various additives in the culture medium, specifically without additives for specific differentiation or promoting different tissue growth. Therefore, besides electrolytes, selection agents and medium stabilisers, the culture medium preferably does not contain any further additives for cell supply. The culture medium for stably transformed plants is preferably free from added sugars, phytohormones or mixtures thereof. The culture medium for transiently transformed protoplasts is prefer-

ably free from added phytohormones.

Preferred moss cells are moss cells of the group Physcomitrella, Funaria, Sphagnum, Ceratodon, Marchantia and Sphaerocarpos, especially in protonema cultures.

According to another aspect, the present invention also provides the use of an isolated nucleic acid molecule encoding an MEPR for industrially producing a polypeptide, especially for providing recombinant cells producing said polypeptide. The industrial production allows a large scale preparation of a given polypeptide of interest in bioreactors, e.g. in gram amounts or even higher (commercial yields). This in contrast to the production sufficient for research use (mg amounts) or analytical purposes ( $\mu$ g amounts), which may, of course also be performed by the present invention. In transient expression systems, protein amounts sufficient for such analytical purposes can easily be obtained with the present DNA molecules.

Accordingly, the present invention also encompasses the use of an isolated nucleic acid molecule encoding a MEPR for expression of a moss polypeptide, the expression of said moss polypeptide being not naturally controlled by said MEPR, especially for providing recombinant moss cells expressing said polypeptide. This use may be reduced to practice both, for research purposes and for industrial scale production of moss polypeptides.

According to another aspect, the present invention also provides the use of an isolated nucleic acid molecule encoding a MEPR for expression of proteins involved in specific posttranslational modifications (e.g. glycosyltransferases), especially for providing recombinant moss cells expressing polypeptides with posttranslational modifications normally not existing or normally existing in another ratio in untransformed moss cells.

According to another aspect, the present invention also provides the use of an isolated nucleic acid molecule encoding a MEPR for expression of proteins involved in metabolic pathways, especially for providing recombinant moss cells altered in their contents of metabolites e.g. secondary metabolites.

According to another aspect, the present invention also provides the use of an isolated nucleic acid molecule encoding a MEPR for expression of antisense molecules, siRNA molecules or ribozymes especially for providing recombinant moss cells with reduced amounts of specific proteins resulting in altered phenotypes e.g. morphologically, biochemically.

According to another preferred aspect, the present invention also relates to the use of an isolated nucleic acid molecule encoding an MEPR according to the present invention for recombinant expression of postranslationally modifying proteins, especially for the production of posttranslationally modified proteins. With such a technology, it is possible to produce proteins which are specifically modified postranslationally (differently than in the native host cell, thereby enabling e.g. plant cells or moss cultures to allow the production of proteins with e.g. mammal or even human glycosylation patterns. Examples wherein such techniques are applied with specific glycosyltransferases are described e.g. in WO 00/49153 A and WO 01/64901 A.

Another preferred use of the isolated nucleic acid molecule encoding an MEPR according to the present invention relates to the in vitro expression of recombinant proteins. The technique of in vitro translation allows a more controlled production of the recombinant product without the need to accept the uncertainties being connected with host cells.

Another preferred use of the nucleic acid molecule according to the present invention is their use for recombinant expression of metabolism modifying proteins, e.g. proteins which modify the (posttranslational) modification of a translated amino acid chain (see e.g. Berlin et al, 1994).

The present invention is further illustrated by the following examples and the figures, yet without being restricted thereto.

Figures:

Fig.1  $\beta$ -tubulin genes in *Physcomitrella patens*,

- Fig.2 Analysis of expression promoting regions of  $\beta$ -tubulins in *Physcomitrella patens*,  
Fig.3 Analysis of expression promoting regions of Pptub 1 by transient transformation of rhVEGF constructs,  
Fig.4 Analysis of expression promoting regions of Pptub 2 by transient transformation of rhVEGF constructs,  
Fig.5 Analysis of expression promoting regions of Pptub 3 by transient transformation of rhVEGF constructs,  
Fig.6 Analysis of expression promoting regions of Pptub 4 by transient transformation of rhVEGF constructs,  
Fig.7 Genomic structure of *Physcomitrella patens* actin genes,  
Fig.8 Comparison of the expression activity of the different 5'actin regions,  
Fig.9 Ppact1 constructs,  
Fig.10 Ppact 5 constructs,  
Fig.11 Ppact 7 constructs,  
Fig.12 Pp act3::vegf constructs,  
Fig.13 Ppact1 promoter:5' intron substitutions,  
Fig.14 Ppact1 promoter:vegf deletion constructs,  
Fig.15 Ppact3 promoter:vegf deletion constructs,  
Fig.16 Ppact5 promoter:vegf deletion constructs,  
Fig.17 Ppact7 promoter:vegf deletion constructs,  
Fig.18 Actin genes in various moss species, and  
Fig.19 Comparison of promoter sequences of homologous actin genes from *Physcomitrella patens* and *Funaria hygrometrica*

## **Material and Methods**

### **Plant material**

*Physcomitrella patens* (Hedw.) B.S.G. has been characterised previously (Reski et al. 1994)). It is a subculture of strain 16/14 which was collected by H.L.K. Whitehouse in Gransden Wood, Huntingdonshire, UK and was propagated by Engel (1968; Am J Bot 55, 438-446).

### **Standard culture conditions**

Plants were grown axenically under sterile conditions in plain inorganic liquid modified Knop medium (1000 mg/l  $\text{Ca}(\text{NO}_3)_2 \times 4 \text{H}_2\text{O}$



250 mg/l KCl, 250 mg/l  $\text{KH}_2\text{PO}_4$ , 250 mg/l  $\text{MgSO}_4 \times 7 \text{ H}_2\text{O}$  and 12.5 mg/l  $\text{FeSO}_4 \times 7 \text{ H}_2\text{O}$ ; pH 5.8 (Reski and Abel (1985) Planta 165, 354-358). Plants were grown in 500 ml Erlenmeyer flasks containing 200 ml of culture medium or on 9 cm Petri dishes with solidified Knop medium (10g/l agar). Flasks were shaken on a Certomat R shaker (B.Braun Biotech International, Germany) set at 120 rpm. Conditions in the growth chamber were  $25 \pm 3^\circ\text{C}$  and a light-dark regime of 16:8 h. Cultures were illuminated from above by two fluorescent tubes (Osram L 58 W/25) providing  $35 \text{ micromol/m}^2\text{s}^{-1}$ . Subculturing of liquid cultures was done once a week by disintegration using an Ultra-Turrax homogenizer (IKA, Staufen, Germany) and inoculation of two new 500 ml Erlenmeyer flasks containing 100 ml fresh Knop medium. Additionally, cultures were filtered 3 or 4 days after disintegration and were transferred into fresh Knop medium.

Bioreactor cultures were grown in Knop medium or in 1/10 Knop medium, respectively, in stirred tank glass bioreactors (Aplikon, Schiedam, The Netherlands) with a working volume of 5 liters (as described in Hohe and Reski, Plant Sci. 2002, 163, 69-74). Stirring was performed with a marine impeller running with a speed of 500 rpm, the cultures were aerated with  $0.3 \text{ vvm}$  [(aeration volume)/(medium volume)/min] air. The culture temperature of  $25^\circ\text{C}$  in the vessel was controlled by a double jacket cooling system. Light intensity was  $50 \text{ micromol/m}^2\text{s}^{-1}$  provided by fluorescent tubes (Osram L 8W/25) with a light/dark rhythm of 16/8 h. The pH-value in the cultures (pH 6.5 - 7.0) was not adjusted.

### **Protoplast Isolation**

Different protocols for the isolation of protoplasts (Grimsley et al. 1977; Schaefer et al. 1991; Rother et al. 1994; Zeidler et al. 1999; Hohe and Reski 2002; Schaefer 2001) have been described for *Physcomitrella patens*. For the work presented herein, a modification/combination of the previously described methods was used:

Moss tissue was cultivated for 7 days in Knop medium with reduced (10%)  $\text{Ca}(\text{NO}_3)_2$  content. Cultures were filtered 3 or 4 days

after disintegration and were transferred into fresh Knop medium with reduced (10%)  $\text{Ca}(\text{NO}_3)_2$  content. After filtration the moss protonemata were preincubated in 0.5 M mannitol. After 30 min, 4% Driselase (Sigma, Deisenhofen, Germany) was added to the suspension. Driselase was dissolved in 0.5 M mannitol (pH 5.6-5.8), centrifuged at 3600 rpm for 10 min and sterilised by passage through a 0.22  $\mu\text{m}$  filter (Millex GP, Millipore Corporation, USA). The suspension, containing 1% Driselase (final concentration), was incubated in the dark at RT and agitated gently (best yields of protoplasts were achieved after 2 hours of incubation). The suspension was passed through sieves (Wilson, CLF, Germany) with pore sizes of 100 micrometer and 50 micrometer. The suspension was centrifuged in sterile centrifuge tubes and protoplasts were sedimented at RT for 10 min at 55 g (acceleration of 3; slow down at 3; Multifuge 3 S-R, Kendro, Germany). Protoplasts were gently resuspended in W5 medium (125 mM  $\text{CaCl}_2 \times 2 \text{ H}_2\text{O}$ ; 137 mM NaCl; 5.5 mM glucose; 10 mM KCl; pH 5.6; 660-680 mOsm; sterile filtered; Menczel et al. 1981). The suspension was centrifuged again at RT for 10 min at 55 g (acceleration of 3; slow down at 3; Multifuge 3 S-R, Kendro, Germany). Protoplasts were gently resuspended in W5 medium. For counting protoplasts a small volume of the suspension was transferred to a Fuchs-Rosenthal-chamber.

### **Transient Transformation**

Different protocols for transformation (Schaefer et al. 1991; Reutter and Reski 1996, Schaefer 2001) have been described for *Physcomitrella patens*. For the work presented herein, a modification/combination of the previously described methods was used:

For transformation protoplasts were incubated on ice in the dark for 30 minutes. Subsequently, protoplasts were sedimented by centrifugation at RT for 10 min at 55 g (acceleration of 3; slow down at 3; Multifuge 3 S-R, Kendro). Protoplasts were resuspended in 3M medium (15 mM  $\text{CaCl}_2 \times 2 \text{ H}_2\text{O}$ ; 0.1% MES; 0.48 M mannitol; pH 5.6; 540 mOsm; sterile filtered, Schaefer et al. (1991) Mol Gen Genet 226, 418-424) at a concentration of  $1.2 \times 10^6$  protoplasts/ml. 250 microliter of this protoplast suspension were dispensed into a new sterile centrifuge tube, 50 microliter DNA

solution (column purified DNA in H<sub>2</sub>O (Qiagen, Hilden, Germany, Hilden, Germany); 10-100 microliter optimal DNA amount of 60 microgram was added and finally 250 microliter PEG-solution (40% PEG 4000; 0.4 M mannitol; 0.1 M Ca(NO<sub>3</sub>)<sub>2</sub>; pH 6 after autoclaving) was added. The suspension was immediately but gently mixed and then incubated for 6 min at RT with occasional gentle mixing. The suspension was diluted progressively by adding 1, 2, 3 and 4 ml of 3M medium. The suspension was centrifuged at 20°C for 10 minutes at 55 g (acceleration of 3; slow down at 3; Multifuge 3 S-R, Kendro). The pellet was resuspended in 400 microliters 3M medium. Cultivation of transformed protoplasts was performed in 48 well plates (Cellstar, greiner bio-one, Frickenhausen, Germany).

Transient transformations were incubated in dim light (4.6 micromols-lm<sup>-2</sup>) at 25°C. Samples were taken after 24h and 48h, respectively, by carefully replacing half of the medium (200 microliters) by fresh medium. The medium was not replaced completely since the protoplasts have to be kept in liquid. The removed medium (including recombinant protein) was stored at -20°C. The 48h samples were measured in an ELISA.

### **Stable transformation**

Different protocols for transformation (Schaefer et al. 1991; Reutter and Reski 1996, Protocol Schaefer 2001) have been described for *Physcomitrella patens*. For the work presented herein, a modification/combination of the previously described methods was used:

For transformation protoplasts were incubated on ice in the dark for 30 minutes. Subsequently, protoplasts were sedimented by centrifugation at RT for 10 min at 55 g (acceleration of 3; slow down at 3; Multifuge 3 S-R, Kendro). Protoplasts were resuspended in 3M medium (15 mM CaCl<sub>2</sub> x 2 H<sub>2</sub>O; 0.1% MES; 0.48 M mannitol; pH 5.6; 540 mOsm; sterile filtered, Schaefer et al. (1991) Mol Gen Genet 226, 418-424) at a concentration of  $1.2 \times 10^6$  protoplasts/ml. 250 microliter of this protoplast suspension were dispensed into a new sterile centrifuge tube, 50 microliter DNA solution (column purified DNA in H<sub>2</sub>O (Qiagen, Hilden, Germany,

Hilden, Germany); 10-100 microliter optimal DNA amount of 60 microgram was added and finally 250 microliter PEG-solution (40% PEG 4000; 0.4 M mannitol; 0.1 M  $\text{Ca}(\text{NO}_3)_2$ ; pH 6 after autoclaving) was added. The suspension was immediately but gently mixed and then incubated for 6 min at RT with occasional gentle mixing. The suspension was diluted progressively by adding 1, 2, 3 and 4 ml of 3M medium. The suspension was centrifuged at 20°C for 10 minutes at 55 g (acceleration of 3; slow down at 3; Multifuge 3 S-R, Kendro). The pellet was re-suspended in 3 ml regeneration medium. Selection procedure was performed as described by Strepp et al. (1998).

#### **ELISA**

Recombinant VEGF121 expressed by transient transformed moss protoplasts was quantified by ELISA (R&D Systems, Wiesbaden, Germany). The ELISA was performed according to the instructions of the manufacturer. The samples were diluted for quantification.

#### **Bacterial strains and cloning vectors**

For all cloning and propagation experiments *Escherichia coli* strain Top10 (Invitrogen, Karlsruhe, Germany) was used. For cloning of DNA-fragments pCR2.1-TOPO (Invitrogen, Karlsruhe, Germany), pCR4-TOPO (Invitrogen, Karlsruhe, Germany), pZerO-2 (Invitrogen, Karlsruhe, Germany) or pRT101 (Töpfer et al. (1987), NAR, 15, p5890) were used as vectors.

#### **Genomic DNA: preparation, digestion, ligation**

*Physcomitrella patens* genomic DNA was isolated from 13 days old protonemata following the CTAB protocol (Schlink and Reski, 2002).

Genomic DNA (3-5 micrograms) was digested with 30 units of various restriction endonucleases (e.g. BamHI, EcoRI, HindIII, KpnI, NcoI, NdeI, PaeI, PstI, XbaI; all MBI Fermentas, St. Leon-Rot, Germany) in a total volume of 30 microliters for two hours at 37°C, using one endonuclease per digest. Digested DNA was purified using PCR Purification Columns (Qiagen, Hilden, Germany),

following the suppliers manual (30 microliters digest + 200 microliters buffer PB). Elution was done in 50 microliters Elution Buffer (EB; Qiagen, Hilden, Germany). Prior further treatment, 10 microliters of the eluate were analysed on an agarose gel (0,5%).

The remaining DNA was religated with 5 units T4 Ligase (MBI Fermentas, St. Leon-Rot, Germany) in a total volume of 300 microliters for two hours at RT and additional two days at 4°C. Prior addition of the enzyme ligation mixtures were put for five minutes at 50°C and then on ice, in order to melt sticky end basepairing. After ethanol precipitation with 0,3 M Na-acetat (pH 4.8) and two washes with 70% ethanol the religated DNA was resuspended in 200 microliters EB. One to three microliters of this religated genomic DNA were used for I-PCR.

#### **RNA Preparation**

*Physcomitrella patens* total RNA was prepared by grinding tissue under liquid nitrogen and by the usage of E.Z.N.A. Plant RNA Kit (PeqLab) or RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the suppliers manuals. Total RNAs were gel analysed, quantified (OD260), and stored at -20°C or -80°C, respectively.

#### **DNase treatment and First Strand cDNA Synthesis**

1 microgram of total RNAs was DNase (GIBCO BRL) digested in a total volume of 11 microliters, following the suppliers manual. 4,5 microliters of this DNase treated total RNA (~400ng) was used with Oligo dT(12-18) primers and SUPERScript II RNase H Reverse Transcriptase (GIBCO BRL) to prepare first strand cDNA, following the suppliers manual. The resulting cDNA was 10 times diluted with sterile ddH<sub>2</sub>O and stored at -20°C.

#### **PCR in general**

If not indicated in particular PCRs were done with Advantage cDNA Polymerase Mix (BD Biosciences Clontech, Heidelberg, Germany). For all other PCR-approaches the following DNA polymerases were used: Taq recombinant polymerase (MBI Fermentas,

St. Leon-Rot, Germany), Pfu native polymerase (MBI Fermentas, St. Leon-Rot, Germany), Platinum Pfx DNA polymerase (Invitrogen, Karlsruhe, Germany) or TripleMaster PCR System (Eppendorf, Hamburg, Germany). Licenced Thermo-cyclers were Mastercycler gradient (Eppendorf, Hamburg, Germany). All primers were synthesised by MWG Biotech AG, Ebersberg, Germany. For PCR product purification or gel elution GFX PCR DNA and Gel Band Purification Kit (Amersham Bioscience, Freiburg, Germany) was used, following the suppliers manual.

### **Construction and Cloning of Recombinant Plasmids**

Conventional molecular biology protocols were essentially as described by Sambrook et al. (1989), Molecular Cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

### **Inverse PCR (I-PCR) & nested PCR**

I-PCR was done with 0.25 microliters Advantage cDNA Polymerase Mix and buffer (including 3,5 mM Mg(OAc)<sub>2</sub>, both BD Biosciences Clontech, Heidelberg, Germany), 0.2 mM each primer, 0.2 mM dNTPs and one to three microliters of genomic religations (see above) in a total volume of 25 microliters. Cycling conditions were: an initial step of 2 minutes at 96°C, then 20 seconds 96°C, 10 seconds initially 67°C (touchdown: -0.15°C/cycle) and 10 minutes 68°C as a second step, with 35 to 40 repetitions, followed by a terminal step of 20 minutes at 68°C and cooling to 4°C at the end of the program. PCR products were eluted from agarose gels. Elution was done in 30 microliters. Eluted PCR products were either cloned directly in TOPO TA vectors (pCR4-TOPO, Invitrogen, Karlsruhe, Germany) or used as template for reconfirmation in nested PCRs. In the latter case gel eluted, nested PCR products were cloned in TOPO TA vectors (pCR4-TOPO, Invitrogen, Karlsruhe, Germany). Cycling conditions for nested PCRs were: an initial step of 1 minutes at 96°C, then 20 seconds 94°C, 10 seconds 56°C and 4 minutes 68°C as a second step, with 25 repetitions, followed by a terminal step of 10 minutes at 68°C.

### **Generation of pRT101new for cloning of amplified promoter frag-**

## **ments**

pRT101p21 (Gorr 1999) was reamplified with Pfu native polymerase (MBI Fermentas, St. Leon-Rot, Germany) using primer 320 and 321 (for this and all subsequent primers see Table 1). Primer 320 (forward) starts at the 2nd codon (5'-(atg)aac...) of the VEGF signal peptide. Primer 321 (reverse) starts in the middle of the HincII site within the multiple cloning site in front of the 35S promoter (5'-gac...). An additional XhoI site was introduced with primer 321. Religation of the PCR product resulted in loss of the 35S promoter and the reconstitution of a HincII site. The sequence of the VEGF gene was verified by sequencing. This new vector was called pRT101new and used for cloning of expression promoting regions via the XhoI or HincII site, respectively, in front of the reporter gene.

## **Sequencing**

All sequencing reactions were performed by SEQLAB Sequence Laboratories, Göttingen, Germany

## **Software**

Sci Ed Central, Clone Manager Suite were used for primer design, pairwise and multiple sequence alignments. Lasergene, DNASTAR (Version 5) Megalign and SeqMan was used for analysing sequencing data. Homology searches were carried out by BLAST 2 (Altschul et al., 1997).

## **EXAMPLES**

The present invention is illustrated by four examples for moss expression promoting regions: first, the isolation and analysis of various members of a family of tubulin expression promoting regions of *Physcomitrella patens*. In the second example expression promoting regions for the actin gene family from a variety of different mosses are provided. The third and fourth example deals with ubiquitin expression promoting regions and with RBCS expression promoting regions.

**EXAMPLE 1: Cloning and analysis of *Physcomitrella patens*  $\beta$ -tubulin genes and their expression promoting regions.**

**Overview**

In order to get  $\beta$ -tubulin (tub) regulatory/promoter sequences from *Physcomitrella patens* (Pp) in a first step coding sequences of  $\beta$ -tubulin homologues were isolated by polymerase chain reaction (PCR). Therefor an alignment of all nine published  $\beta$ -tubulin genomic sequences from *Arabidopsis thaliana* (Attub 1-9) were used to design primers within highly conserved coding regions (8F, 9F and 10R; for this and all subsequent primers see Table 1). In addition, sequence information of public EST data from *Physcomitrella patens* were used, but only three did show homologies to  $\beta$ -tubulins. One of which was used to design a gene-specific primer (F7) upstream of the predicted coding region. Sequence comparison of all cloned PCR products, generated with the primers mentioned and EST data lead to 3 groups of clones with identical DNA within but differences between groups, mainly, but not exclusively, due to differences within introns. This  $\beta$ -tubulin orthologues were named Pptub 1, Pptub 2 and Pptub 3, respectively.

Furthermore, since during the running project, more EST data were available (more than 50000 new entries in NCBI/dbEST with beginning of 2002), a detailed analysis of all 121 *Physcomitrella patens* ESTs with high similarity to  $\beta$ -tubulin lead to three additional new upstream and three downstream groups of ESTs, being identical within a group but neither identical to any other group nor to Pptub 1-3. PCR with primers derived from predicted noncoding upstream and downstream regions (see below) from each new group and permuting all primer combinations helped to correlate corresponding upstream and downstream groups to a particular locus, named Pptub 4, Pptub 5 and Pptub 6, respectively. Both, genomic and cDNA amplicates of all three new loci were cloned and sequenced, raising the number of  $\beta$ -tubulin orthologues in *Physcomitrella patens* to six.

Pptub 1 to 4 (in contrast to Pptub 5 and 6) are much more frequently represented in EST databases. Corresponding cDNA libraries were produced using RNA mainly from protonema and young



gametophore. So, for this four genes only, based on the gained sequence data, an inverse PCR approach (I-PCR) was performed in order to walk into flanking genomic regions.

### **Pptub 1**

As already mentioned in a first step, Taq (MBI Fermentas, St. Leon-Rot, Germany) PCR fragments from two independent PCRs on *Physcomitrella patens* genomic DNA using primers 8F and 10R were cloned. One clone (2-1) and two clones (8-1, 8-2), respectively, from each PCR were sequenced partially and turned out to be identical. The corresponding locus was named Pptub 1.

This preliminary sequence information was used to design primers in order to perform a genomic walk into flanking regions of Pptub 1, using an I-PCR approach on religated EcoR I and Hind III genomic digests (primers 35, 36). Reconfirmation of products was done by nested PCR (primers 40, 38). Two clones generated by nested PCR products (E#1 and H 1.7) were sequenced completely.

The Hind III clone H 1.7 did not harbour an internal Hind III site, most likely due to star activity of the enzyme or ligation of a random ds breakage. However, sequences upstream of the first EcoR I site were confirmed by two independent PCRs on genomic DNA (primers 113, 67 and 113, 90). In addition, an additional cDNA (89, 91; Pfu native (MBI Fermentas, St. Leon-Rot, Germany)) PCR product was cloned.

All mentioned clones helped to generate and reconfirm sequence data. In total ~1500 bp upstream of the startcodon and ~1500 bp downstream of the stopcodon were gained.

### **Pptub 2**

As already described above sequence information of published ESTs from *Physcomitrella patens* was used to design a gene-specific primer (F7) upstream of the predicted coding region. PCR on *Physcomitrella patens* genomic DNA (primers F7, 10R) and subsequent cloning and sequencing of the PCR product proofed that it, together with all three so far published Pptub ESTs (Pptub EST 1-3) belong to one locus, named Pptub 2. Intron positions could be verified by comparing EST with genomic sequences.

This preliminary sequence information was used to design gene-specific primers within introns (primers 95 and 71) in order to perform a genomic walk into adjacent genomic regions of Pptub 2, using an I-PCR approach on religated Pag I, BamH I and Nde I genomic digests. PCR products were reconfirmed by nested PCR (primers 38, 35). Two clones generated by nested PCR products (C#2Pag and D#2Nde) were sequenced completely. The Nde I clone D#2 did not harbour an internal Nde I site, most likely due to star activity of the enzyme or ligation of a random ds breakage. However, sequence data were confirmed by C#2Pag and a third I-PCR clone (95#8BamHI; primer 149 and 71). In addition two independent PCRs on genomic DNA (primers 205, 149; Taq (MBI Fermentas, St. Leon-Rot, Germany) and primers 205, 206) confirmed product length. The 205-206 PCR product and an additional genomic downstream PCR product (primers 71, 206; Pfu native (MBI Fermentas, St. Leon-Rot, Germany)) were cloned and helped to verify sequence data.

All mentioned clones helped to generate and reconfirm sequence data. In total ~1400 bp upstream of the startcodon and ~1400 bp downstream of the stopcodon were gained.

### **Pptub 3**

As already mentioned in a first step, Taq (MBI Fermentas, St. Leon-Rot, Germany) PCR fragments from two independent PCRs on *Physcomitrella patens* genomic and cDNA using primers 9F and 10R were cloned. Clones from each PCR (#3-3 genomic, #4-3 cDNA) were sequenced partially and turned out to be identical. The corresponding locus was named Pptub 3.

This preliminary sequence information was used to design gene-specific primers within introns (primers 69, 70) in order to perform a genomic walk into adjacent regions of Pptub 3, using an I-PCR approach on religated Pag I and Nco I genomic digests. Reconfirmation of PCR products was done by nested PCR (primers 38, 35). Two clones (A#1Nco and #4-1Pag) were sequenced completely. A#1Nco is a clone generated by a nested PCR product (38, 35) whereas #4-1PagI was generated by the original I-PCR product (69, 70). In addition a genomic PCR product (primers

203, 204) was cloned and helped to verify sequence data.

All mentioned clones helped to generate and reconfirm sequence data. In total ~1900 bp upstream of the startcodon and ~1100 bp downstream of the stopcodon were gained.

#### **Pptub 4**

As already mentioned, in case of Pptub 4, EST data were used to design gene-specific downstream and upstream primers (297, 299) in order to generate genomic and cDNA clones. Additional genomic clones using native Pfu polymerase (MBI Fermentas, St. Leon-Rot, Germany) helped to verify sequence data.

Primer 297 and 299 were inverted (337, 383) and used to perform a walk into adjacent genomic regions of Pptub 4, using an I-PCR approach on religated Nde I and Nco I genomic digests. Two clones (48#2Nco and A02#3Nde) and additional genomic clones (primers 547 and 374; Advantage cDNA Polymerase Mix (BD Biosciences Clontech, Heidelberg, Germany) and Triple Master (Eppendorf, Hamburg, Germany)) were generated.

All mentioned clones helped to generate and reconfirm sequence data. In total ~2300 bp upstream of the startcodon and ~1100 bp downstream of the stopcodon were gained.

#### **Pptub 5 and 6**

As already mentioned, in case of Pptub 5 and 6, EST data were used to design gene-specific downstream and upstream primers (Pptub 5: 298, 300 and Pptub 6: 296, 336) in order to generate genomic and cDNA clones of each gene. In case of Pptub 5, additional genomic clones using native Pfu polymerase (MBI Fermentas, St. Leon-Rot, Germany) helped to verify sequence data.

All mentioned clones helped to generate and reconfirm sequence data. In total 2031 bp genomic sequence for Pptub 5 and 3161 bp genomic sequence for Pptub 6 were gained.

#### **Cloning strategies**

Preliminary Pptub 1 (2-1, 8-1, 8-1; all genomic) and Pptub 3 (3-3 genomic, 4-3 cDNA) clones were generated with Taq recombinant

polymerase. PCR products were ligated into TOPO TA vectors (pCR4-TOPO, Invitrogen, Karlsruhe, Germany). PCR conditions were: 2.5 unit Taq recombinant polymerase, enzyme buffer, 3.3 mM MgCl<sub>2</sub> (all MBI Fermentas, St. Leon-Rot, Germany), 0.4 mM each primer, 100 nanograms of cDNA or genomic DNA as template in a total volume of 25 microliters. Cycling conditions were: an initial step of 5 minutes at 95°C, then 45 seconds 95°C, 10 seconds 60°C (primer 8F) or 65°C (primer 9F) and 1 minute 72°C as a second step, with 30 to 35 repetitions, followed by a terminal step of 5 minutes at 72°C and cooling to 4°C at the end of the program.

All other genomic and cDNA clones were

Pptub 1: 113-67, 113-90, 89-90, 89-91 cDNA

Pptub 2: F7/R10, 205-206, 71-206

Pptub 3: 203-204

Pptub 4: 547-374 (+ TrippleMaster), 297-299 cDNA + genomic (+ Pfu)

Pptub 5: 298-300 cDNA + genomic (+ Pfu)

Pptub 6: 296-336 cDNA + genomic

Underlined clones above were generated with Advantage cDNA Polymerase Mix, using 0.25 microliters enzyme mix, buffer (including 3,5 mM Mg(OAc)<sub>2</sub>, both BD Biosciences Clontech, Heidelberg, Germany), 0.25 mM each primer, 0.25 mM dNTPs and 10-20 nanograms of template per 20 microliter PCR. Cycling conditions were: an initial step of 2 minutes at 96°C, then 20 seconds 96°C, 10 seconds 60°C and 2 minutes/kb 68°C as a second step, with 35 to 40 repetitions, followed by a terminal step of 15 minutes at 68°C and cooling to 4°C at the end of the program. PCR products of appropriate length were eluted from agarose gels. Elution was done in 30-50 microliters, depending on amount of amplificate. Eluted PCR products were cloned in TOPO TA vectors (pCR4-TOPO, Invitrogen, Karlsruhe, Germany).

All other clones were generated with Pfu native polymerase, as were the two additional genomic clones 297-299 and 298-300, using 0.3 microliters polymerase (= 0.75 units), buffer, 2-4 mM MgSO<sub>4</sub> (all MBI Fermentas, St. Leon-Rot, Germany), 0.25 mM each

primer, 0.2 mM dNTPs and 10-20 nanograms of template per 20 microliter PCR. Cycling conditions were: an initial step of 2 minutes at 96°C, then 20 seconds 96°C, 10 seconds 60°C and 2 minutes/kb 72°C as a second step, with 35 to 40 repetitions, followed by a terminal step of 10 minutes at 72°C and cooling to 4°C at the end of the program. PCR products of appropriate length were eluted from agarose gels. Elution was done in 30-50 microliters, depending on amount of amplificate. Eluted PCR products were cloned in pZErO-2 (Invitrogen, Karlsruhe, Germany) linearised with EcoRV.

An additional clone of 547-374 was generated with the TripleMaster PCR System, using 0.25 microliters polymerase mix (= 1.25 units), tuning buffer (including 2.5 mM Mg<sup>2+</sup>, both Eppendorf, Hamburg, Germany), 0.2 mM each primer, 0.2 mM dNTPs and 10-20 nanograms of template per 20 microliter PCR. Cycling conditions were: an initial step of 2 minutes at 96°C, then 20 seconds 96°C, 20 seconds 60°C and 3 minutes 72°C as a second step, with 40 repetitions, followed by a terminal step of 10 minutes at 72°C and cooling to 4°C at the end of the program. PCR products of appropriate length were eluted from agarose gels. Elution was done in 30-50 microliters, depending on amount of amplificate. Eluted PCR products were cloned in TOPO TA vectors (pCR4-TOPO, Invitrogen, Karlsruhe, Germany).

In summary, PCR on genomic DNA of *Physcomitrella patens* and cloning of PCR products lead to sequence information of six transcribed *Physcomitrella patens*  $\beta$ -tubulin genes. Additionally, EST and cDNA data were used to confirm genomic sequence data and intron/exon borders. In case of Pptub 1 to 4 inverse PCR lead to non transcribed flanking 5' and 3' genomic sequences. A general overview of all six genomic regions is given in Figure 1.

### **Gene structure & Conservation**

As already stressed, Pptub 1 to 4 are most abundantly represented in EST databases. In addition the great majority of their corresponding ESTs were raised from full length cDNA libraries. This two facts helped to determine the transcriptional start site (TSS) of Pptub 1 to 4 *in silico*. A multiple alignment of 5'

ESTs against corresponding upstream genomic regions showed that Pptub 1 to 3 do have a precise transcriptional initiation: 20 out of 27 5' ESTs for Pptub 1, 16 out of 20 5' ESTs for Pptub 2 and 9 out of 14 5' ESTs for Pptub 3, do start at the same, most upstream position, marked with +1 (Figure 3-6). In addition all three TSSs are surrounded by a consensus sequence (see below). In case of Pptub 4 the 23 5' ESTs indicate multiple TSSs within 100 bp. The start site of the most upstream 5' EST was defined as +1.

An analogous multiple alignment of 3' ESTs against corresponding downstream genomic regions reconfirmed that plant genes almost always come with more than one poly(A) site and that consensus sequences are much less sharply defined than in e.g. mammalian genes, in which the sequence AAUAAA is nearly ubiquitous (for review see: Rothnie et al., 1996).

The six cloned loci of *Physcomitrella patens* did not show any nonsense stop-codons and proper proteins with high similarities to known  $\beta$ -tubulins could be predicted. Outside the coding regions generally, the similarity drops immediately and significantly. Concerning 5' putative regulatory elements, a detailed comparison of all four upstream regions revealed no overall conservation within the gene family or to 5' regions of other known plant  $\beta$ -tubulin genes. However, some interesting matches of conservation within the gene family could be detected:

- a) The determined TSSs of Pptub 1 to 3 in all three cases fall within the consensus sequence T/C C A(+1) G/C T G T G C and are embedded in C/T-rich regions (compare consensus of 171 unrelated TATA plant promoters: T/C C A(+1) N M N in plantProm Database under <http://mendel.cs.rhul.ac.uk/mendel.php?topic=plantprom>).
- b) 22-24 bp upstream of the TSS -which is within the typical distance for plant TATA promoters (see plantProm DB)- a weak 8 bp TATA box embedded in a conserved stretch of 20-25 bp can be found in Pptub 1 to 3. The TATA box consensus from 171 unrelated plant promoters is: T<sub>96</sub> A<sub>95</sub> T<sub>96</sub> A<sub>100</sub> A<sub>62</sub>/T<sub>38</sub> A<sub>97</sub> T<sub>61</sub>/A<sub>38</sub> A<sub>73</sub> (see plantProm DB) and for Pptub 1-3 is: T t T A T c T c/t/A, with capitals indicating correlation to consensus.
- c) all four genes do have a very low degree of Adenosine (9-

16%) in their 5'UTRs.

d) The 5' UTR of Pptub 4 has an overall C/T content of 74%, which -in addition- harbours a C/T stretch (~ 50 bp), directly behind the start point of the shortest, most downstream 5' EST.

e) Pptub 2 harbours a 40 bp polyA stretch around 450 bp upstream of the TSS (-450 until -489).

f) In Pptub 1 and 4 upstream of app. position -420 long very A/T-rich regions begin (Pptub 1 over 80% A/T for nearly 900 bp and Pptub 4 75% A/T for 1750 bp, rendering open the possibility for the location scaffold/matrix attached regions (S/MARs; (Liebich et al., 2002) upstream of this genes.

### **Functional Characterization & Quantification of $\beta$ -tubulin promoters**

Definition of minimal promoter-fragments giving a maximum of promoter activity was done by functional quantification of putative 5' regulatory sequences of Pptub 1 to 4 in a transient expression system, using nonregenerating *Physcomitrella patens* protoplasts as expression system. For each promoter several constructs of different lengths including upstream regions and 5' UTRs, were brought precisely in front of the startcodon of the reporter gene. As reporter gene a human protein (recombinant human vascular endothelial growth factor 121: rhVEGF121; Gorr 1999) was secreted into the medium via its own signalpeptide. The amount of rhVEGF121 in the supernatant of the moss culture was quantified by an ELISA and reflected the strength of the promoter or promoter fragment in the system. Values were related to values obtained by the 35S promoter. Each construct was transformed a minimum of six times in two to three different transformation experiments. Samples were taken after 24 and 48 hours, respectively, with 48 hour samples measured twice in appropriate dilutions in an ELISA. An overview of the results is given in Figure 2.

The expression promoting regions of Pptub 1 to 4 are disclosed as Seq.ID.Nos. 1 to 8.

**Cloning of amplified promoter fragments of Pptub 1 and 4 into pRT101new**

Pptub 1: 1-0 (primer 364XhoI, 363cat)  
1-1 (primer 219XhoI, 363cat)  
1-3 (primer 549XhoI, 363cat)  
1-4 (primer 226XhoI, 363cat)  
1-5 (primer 550XhoI, 363cat)

Pptub 2: 2-0 (primer 291, 225cat)

Pptub 3: 3-0 (primer 292, 223cat)

Pptub 4: 4-0 (primer 373XhoI, 374cat)  
4-1 (primer 548XhoI, 374cat)

The promoter fragments given above were amplified with Pfu native polymerase (MBI Fermentas, St. Leon-Rot, Germany) on genomic DNA using reverse primers starting with the reverse complement sequence of the ATG start codon (cat...) and, in part, forward primers containing XhoI sites. PCR products were cut XhoI and ligated into XhoI/HincII or not cut at all and and ligated into HincII opened pRT101new, respectively. Generated clones were verified by sequencing. Clone 1-2 (XhoI/EcoRI), 2-1 (BglII), 2-2 (SalI), 2-3 (EcoRI/SalI), 2-4 (EcoRI/SalI), 3-2 (SalI), 3-3 (Eco147I/HincII), 3-4 (XhoI/SalI) were generated by internal deletions of longer clones. The remaining vectors were gel-eluted and religated. In case single strand overhangs did not fit, ligation was performed after filling-in of recessed 3'-termini with Klenow Fragment (MBI Fermentas, St. Leon-Rot, Germany), following the suppliers manual.

#### **Pptub 1**

Six different promoter lengths were cloned into the transformation vector pRT101p21 in front of the reporter gene. The data of all constructs are given in figure 3. (5' UTR = +1 (TSS) until +226, +227= ATG)

1-0	-1307 bp	(1533 bp 5' region of Pptub 1)
1-1	- 985 bp	(1211 bp 5' region of Pptub 1)
1-2	- 416 bp	(642 bp 5' region of Pptub 1)
1-3	- 248 bp	(474 bp 5' region of Pptub 1)



- 1-4 - 83 bp (309 bp 5' region of Pptub 1)
- 1-5 - 71 bp (297 bp 5' region of Pptub 1)

Promoter fragment 1-2 can be defined as the shortest promoter fragment giving high expression rates. The rates are app. 150% compared to values generated with the 35S promoter, which was set to 100%. Note that upstream of the minimal promoter fragment 1-2 a long, very A/T rich region starts (over 80% A/T for nearly 900 bp).

### **Pptub 2**

Five different promoter lengths were cloned into the transformation vector pRT101p21 in front of the reporter gene. The data of all constructs are given in Figure 4. (5' UTR = +1 (TSS) until +122, +123= ATG)

- 2-0 -1075 bp (1197 bp 5' region of Pptub 2)
- 2-1 - 676 bp (798 bp 5' region of Pptub 2)
- 2-2 - 425 bp (547 bp 5' region of Pptub 2)
- 2-3 - 245 bp (367 bp 5' region of Pptub 2)
- 2-4 - 67 bp (189 bp 5' region of Pptub 2)

Promoter fragment 2-2 can be defined as the shortest promoter fragment giving high expression rates. The rates are comparable to values generated with the 35S promoter (100%).

### **Pptub 3**

Different promoter lengths were cloned into the transformation vector pRT101p21 in front of the reporter gene. The data of four constructs are given in Figure 5. (5' UTR = +1 (TSS) until +112, +113= ATG)

- 3-0 -1274 bp (1386 bp 5' region of Pptub 3)
- 3-2 - 765 bp (879 bp 5' region of Pptub 3)
- 3-3 - 272 bp (384 bp 5' region of Pptub 3)
- 3-4 + 52 bp (60 bp 5' UTR of Pptub 3)

Promoter fragment 3-2 can be defined as the shortest promoter fragment giving high expression rates. The rates are app. 300% compared to values generated with the 35S promoter, which was

set to 100%.

#### **Pptub 4**

Two different promoter lengths were cloned into the transformation vector pRT101p21 in front of the reporter gene. The data are given in Figure 6.

(5' UTR = TSSs (+1 until +103) until +205, +206= ATG)

4-0	-419 bp	(624 bp 5' region of Pptub 4)
4-1	- 1 bp	(206 bp 5' region of Pptub 4)

Promoter fragment 4-1 gives expression rates that are app. 250% compared to values generated with the 35S promoter, which was set to 100%. Note that upstream of this minimal promoter fragment (4-0) a long, very A/T rich region starts (75% A/T for 1750 bp).

In summary transient promoter activity of Pptub 1 to 4 genomic upstream regions were characterised. Minimal promoter fragments showing a maximum of promoter activity were defined and gave yields of up to 3 times the 35S promoter activity.

Pptub-constructs summary (see also: sequence listing)

#### **Pptub1 upstream**

- 1533 until -1 (+1 = start codon)
- 1533 until -644 = 81 % AT
- 1533 VEGF 1-0 (primer 364)
- 1211 VEGF 1-1 (primer 219)
- 642 VEGF 1-2 (EcoRI/XhoI)
- 474 VEGF 1-3 (primer 549)
- 309 VEGF 1-4 (primer 226)
- 297 VEGF 1-5 (primer 550; without putative TATA box: -304 until -295)
- 226 TSS (start of 5'UTR)

#### **Pptub1 downstream**

- 1 until 1539 (1 = directly behind stop codon)
- 332 end of longest EST (3'UTR)
- 1539 start of primer 90

Pptub2 upstream

-1197 until -1 (+1 = start codon)  
-1197 VEGF 2-0 (primer 291)  
-798 VEGF 2-1 (BglII)  
-547 VEGF 2-2 (SalI)  
-450 until -489 = poly A stretch  
-367 VEGF 2-3 (EcoRI/SalI)  
-189 VEGF 2-4 (XhoI/SalI)  
-122 TSS (start of 5'UTR)

Pptub2 downstream

1 until 1012 (1 = directly behind stop codon)  
297 end of longest EST (3'UTR)  
1012 start of primer 206

Pptub3 upstream

-1386 until -1 (+1 = start codon)  
-1386 VEGF 3-0 (primer 292)  
-879 VEGF 3-2 (SalI)  
-384 VEGF 3-3 (Eco147I/HincII)  
-112 TSS (start of 5'UTR)  
-60 VEGF 2-4 (XhoI/SalI)

Pptub3 downstream

1 until 997 (1 = directly behind stop codon)  
203 end of longest EST (3'UTR)  
1012 start of primer 204

Pptub4 upstream

-624 until -1 (+1 = start codon)  
-624 VEGF 4-1 (primer 373)  
-206 VEGF 4-2 (primer 548)  
-205 until -103 area of TSS (start of 5'UTR)  
-55 until -93 CT stretch

Pptub4 downstream

1 until 1146 (1 = directly behind stop codon)  
466 end of longest EST (3'UTR)  
1141 until 1164 NcoI

**EXAMPLE 2: Cloning and analysis of actin genes from different moss species and their expression promoting regions.**

**2.1. Genomic structure of *Physcomitrella patens* actin genes.**

Four actin genes and promoter regions of the moss *Physcomitrella patens* and three from *Funaria hygrometrica* and the liverwort *Marchantia polymorpha* have been isolated in order to construct expression vectors for their use in moss.

Using specific oligos designed from *Physcomitrella* EST sequences that are present in the public databases, four actin genes (Ppact1, Ppact3, Ppact5 and Ppact7) were isolated in several rounds of iPCR from genomic DNA and sequenced.

In *Physcomitrella* the structure of the isolated genes resembles in one case (Ppact1) the conserved structural organisation of actin genes of higher plants. The un-translated leader is disrupted by a relatively long (955bp) intron located 14 nt upstream the initiator ATG. The coding region presents three smaller introns which are situated at the same positions as the introns of actin genes of other plant species. The first one is located between codons 20 (lys) and 21 (ala), the second is splitting codon 152 (gly) and the third is between codon 356 (gln) and 357 (met). This general structure appear to be different for the three other *Physcomitrella* actin genes isolated (Ppact3, Ppact5, and Ppact7). In those cases the 5'UTR intron (434bp, 1006bp and 1055bp respectively) is also located 14nt before the ATG but the coding region is disrupted only by one intron positioned between codons 21 (lys) and 22 (ala) (Fig. 7).

**2.2. Activity studies of the expression promoting regions of actin genes.**

To study the activity of the different *Physcomitrella* actin expression promoting regions (Seq.ID Nos. 5 to 8) as well as the effect of the 5'UTR of the different genes, different vectors were designed for expression of the hVEGF protein under the control of the 5' regions under study.

Around 2kb genomic regions upstream the transcription initiation site were isolated by iPCR from genomic DNA and sequenced, and vectors containing the cDNA of the human VEGF driven by the promoters and containing the exact leader sequences including the 5'intron were constructed for transient transfection of moss protoplasts. The complete 5'promoting expression regions were amplified by proof reading PCR using primer 395 and 332 for Ppact1, 408 and 333 for Ppact3, 511 and 334 for Ppact5, and 413 and 335 for Ppact7.

Transformation of protoplasts was performed using the same number of molecules for each construct to be tested and in parallel to a construct carrying the hVEGF cDNA under the control of the CaMV 35S promoter. The hVEGF protein contains at the N-terminal part a 26 aa signal peptide that permits secretion of the recombinant protein to the medium. Analysis of the transformations was carried out by ELISA, taking different dilutions of the medium where the protoplasts were incubated 48 hours after transformation.

The capacity to drive expression of the different *Physcomitrella* 5' actin regions was compared to the activity of the constitutive 35S promoter.

In all cases analysed, the 5'regions of the actin genes were reaching higher activity than the 35S promoter. However the level of expression varied for the different actin regulatory sequences. Thus, the 5'sequence of Ppact3 was only promoting around a 2 fold higher expression of VEGF than the 35S promoter. Higher levels of VEGF were measured when vectors containing the 5'regions of Ppact1 and Ppact7 were used for transformation. In those cases values between 4 and 8 folds the 35S values were obtained. Nevertheless the most dramatic differences were observed in the case of the 5'Ppact5 gene, where up to 11 fold higher expression values compared to the 35S were in some cases obtained (Fig. 8).

To further investigate on the role of the 5' UTR region of the high activity *Physcomitrella* actin genes, vectors containing de-

letions, combinations and substitutions of the 5'UTR intron were made and used for transient assays in moss protoplasts.

Deletion of the Ppact1 5'intron dramatically decreased the levels of transient expression in comparison to those obtained when the intact 5'region of Ppact1 was used. In this case the amount of secreted VEGF protein that could be detected in the protoplasts medium was very similar to the obtained by the CaMV 35S promoter. This would indicate that the 5'intron of the Ppact1 is essential for efficient gene expression from the Ppact1 promoter. Same results were obtained when the 5'UTR including the leader intron was fused downstream the 35S promoter. This construct yielded the same amount of secreted protein as the intact 35S promoter indicating that the 5'UTR region is not having any dramatic influence on the activity of promoters other than the Ppact1 promoter. It is important to indicate that a construct carrying just the 5'UTR Ppact1 region was able to promote protein production only in a 30% lower amount than the 35S promoter alone. This could suggest a small promoter activity in this region of the gene, or a rest of promoter activity present in the backbone sequence of the vector (Fig. 9).

The same approach was used to investigate the influence on the promoter activities of the 5'UTR introns contained in the Ppact5 and Ppact7 genes. Constructs in which the 5' intron was deleted were analysed and similar results as in the case of Ppact1 were obtained, ie. the amount of protein reached was approximately the same as with the 35S promoter in the case of Ppact5 and slightly lower in the case of Ppact7, indicating that the presence of the intron in the 5'UTR is essential for the efficient activity of the promoters. Again some residual promoting activity was observed when the transformation was performed with constructs containing only the 5'transcribed region up to the ATG. Furthermore, in the case of these two genes, the fusion of the 5'UTR downstream the 35S promoter yielded higher rates (2 to 7 folds) of expression of the VEGF protein when compared to the 35S promoter alone (Fig. 10, 11). Similar results were observed in the case of Ppact3, where the 5'UTR alone or fused downstream the CaMV 35S, yielded around 2 and 3 folds respectively in comparison to the 35S (Fig. 12). These indications would suggest

the presence of enhancer activity in the 5' transcribed regions for these three genes even when they are positioned under a different promoter.

To further investigate the role of the 5' intron present in the Ppact1, Ppact5 and Ppact7 genes, substitutions of the leader intron of the Ppact1 gene with the 5' intron of Ppact5 and Ppact7 were engineered in vectors for transient transformation. In parallel substitutions of the Ppact1 5' intron with the pppact1 introns present in the coding region of the gene, were performed.

Substitutions of the Ppact1 5' intron, by the Ppact 1 coding region introns 1 and 3 resulted in a decrease of the expression levels of around 25%. Still the amount of protein detected was around 2-3 fold higher than the obtained with the CaMV 35S promoter. The substitution of the 5' intron by the intron 2 of the coding region surprisingly resulted in no activity of the promoter (Fig. 13). The construct was however checked, and the sequence showed that the splicing site for the intron was not correct. A new construct carrying the correct splicing sequence was made and the results after moss transformation indicated that the effect of the intron 2 is the same as for the other substitutions.

A reduction of protein expression was also observed when the substitution was done with the 5' introns corresponding to the Ppact5 and Ppact7 genes, but in this case the reduction was slightly smaller.

### **2.3. Deletion constructs of the expression promoting regions of actin genes.**

A further characterisation of the different actin genes promoters was carried out by making deletion constructs of the 5' untranscribed regions and analysing them through transient transformation of moss protoplasts.

Thus for the Ppact1 constructs carrying different genomic region lengths (-1823bp, -992bp, -790bp, -569bp, -383bp, -237bp, and -82bp) upstream the initiation of transcription (+1) were made.

In principle all the constructs except the -82bp, could have full promoter activity. However the -383bp construct shows a reduction of activity and reaches similar levels as the -82bp construct (Fig.14).

Analysis of deletion constructs of the promoter region of Ppact3 revealed some interesting features. As it was described, this promoter presented a lower activity compared to the other actin genes promoters, although in relation to the CaMV 35S, it was slightly more active. In this case the following 5'untranscribed regions were tested: -2210bp, -995bp, -821bp, -523bp, -323bp, -182bp and -81bp. Surprisingly the activity of the promoter was approximately the same as the CaMV 35S for the constructs containing up to -821bp of the promoter region. However the constructs containing from bp -523 and shorter regions towards the transcription start, yielded two folds more amount of recombinant protein. This could indicate cis-acting regions located upstream the -523bp region that down regulate the transcription of this gene during the transient transformation assay (Fig. 15).

In the case of Ppact5, constructs containing the -1872bp, -758bp, -544bp, -355bp, and -121bp fragments upstream the transcription start of the gene were generated. The results obtained from the transient assays indicate that the full activity of the promoter resides in a region between -758 and -121 from the start of transcription (+1) (Fig. 16).

The following deletion constructs for the 5'untranscribed region of Ppact7 were analysed: -1790bp, -1070bp, -854bp, -659bp, -484bp, -299bp, and -66bp. The results obtained indicate that the region comprised in between -484bp and -299bp is essential for the full activity of the promoter during the transient experiment assays. (Fig.17).

In order to obtain a set of heterologous promoters of the *Phycomitrella* actin genes, other two species, the moss *Funaria hygrometrica* and the liverwort *Marchantia polymorpha*, were used to isolate genomic DNA fragments containing actin genes. To this end, oligos with different degrees of degeneration were designed to perform PCR reactions using as template genomic DNA isolated



from the two species.

#### **2.4. Comparison of different actin genes from the different moss species *Physcomitrella patens*, *Funaria hygrometrica* and *Marchantia polymorpha***

##### ***Physcomitrella patens***

The four different genomic actin sequences isolated from *Physcomitrella patens* are likely to represent the whole functional sequences of the genes including 5'promoter sequence, 5'UTR + 5'intron, ORF + internal introns and the 3'UTR and further 3'-downstream sequence. In total for Ppact1 5809 bp, for Ppact3 5633 bp, for Ppact5 8653 bp and for Ppact7 6351 bp of genomic sequence was isolated (Fig. 18 A). The coding regions of the isolated *Physcomitrella* actin cDNAs are almost all 1137 bp in length, except Ppact1 which has an ORF of 1134 bp. The corresponding proteins are 378 amino acids in lengths except Ppact1 which has 377 amino acids. On the nucleotide level the coding sequences share homologies between 86.6 and 98.9 %. The protein sequences have an identity between 97.1 and 99.7 % (DNA STAR, MegAlign Program, Clustal V (weighted) sequence alignment).

For all four *Physcomitrella* actin genes extended genomic DNA sequences 5' of the ATG Start codon could be isolated by iPCR and sequenced: 2973 nt for Ppact1, 3091 nt for Ppact3, 3095 nt for Ppact5 and 3069 nt for Ppact7. For Ppact1, Ppact5 and Ppact7 5'race by using the Gene Racer Kit (Invitrogen), which allows the amplification of only full length cDNAs, was performed to determine the 5'UTRs of the genes. For Ppact3 the 5'UTR was determined by the length of different ESTs from database. By comparing the cDNAs with the genomic iPCR fragments the presence of large 5'introns could be shown. The lengths of the 5'introns which are all located at position -14 to the ATG Start codon are 955 bp, 434 bp, 1006 bp and 1055 bp for Ppact1, Ppact3, Ppact5 and Ppact7 respectively (Fig. 18 A). The positions of the ORF internal introns was determined by comparing the genomic sequences and the derived protein sequences to the cDNA sequences and protein sequences of the actin genes from *Arabidopsis thaliana*. The 5'promoter sequences for the *Physcomitrella* actin genes

available are 1824 nt for Ppact1, 2270 nt for Ppact3, 1909 bp for Ppact5 and 1805 bp for Ppact7 (Fig. 18 A).

In total 4 different actin genes from *Funaria hygrometrica* (expression promoting regions: Seq.ID Nos. 9 to 12) and 3 different genes from *Marchantia polymorpha* (expression promoting regions: Seq.ID Nos. 13 to 15) could be identified by degenerated PCR on genomic DNA. As the aim was predominantly to isolate 5'promoter regions of the putative different actin gene homologs from the different moss species, most of the sequences are incomplete at the 3'end to date (Fig. 18 B/C).

#### *Funaria hygrometrica*

For *Funaria* the identified actin genes were named Fhact1, Fhact4.4, Fhact5 and Fhact5b. 3951 bp of Fhact1, 2417 bp of Fhact4.4, 4432 bp of Fhact5 and 722 bp of Fhact5b of genomic sequence could be isolated by iPCR for the different actin genes. The complete coding cDNA sequence could be isolated for the Fhact1 gene which has a coding sequence of 1134 nucleotides. For the other *Funaria* actin genes partial sequences are available at the moment, lacking the 3'ends: 906 bp for Fhact4.4, 965 bp for Fhact5 and 722 bp for Fhact5b (Fig. 18 B) The isolated coding sequences share homologies in a range of 87.4 and 99.2% on the nucleotide level. The derived protein sequences are 90.8 to 99.2% identical (DNA STAR, MegAlign Program, Clustal V (weighted) sequence alignment).

Except for Fhact5b, 5'sequences upstream of the ATG Start codon could be isolated by iPCR and sequenced. In the case of Fhact1 1824 bp, for Fhact4.4 1333 bp and for Fhact5 3289 bp are available. The length of the different 5'UTRs were determined by 5'race using the Gene Racer Kit (Invitrogen). The intron-exon structure was determined by comparison of the cDNA sequence with the genomic sequences obtained by iPCR and by comparison to the *Physcomitrella* genes. As in the case of the *Physcomitrella* actin genes the identified *Funaria* actin genes contain large 5'introns located at position -14 of the cDNAs, 928 bp, 1015 bp and 656 bp in length for Fhact1, Fhact4.4 and Fhact5 respectively. By now for Fhact1 700 bp, 145 bp for Fhact4.4 and for

Fhact5 2515 bp of 5'promoter sequence was isolated and sequenced. For Fhact1 419 bp of the 3'region was isolated. The 5'regions or 3'regions of the Funaria actin genes are amplified by PCR on genomic DNA from Funaria hygrometrica by using the primers 908 and 909 for the 5' region of Fhact1, 983 and 984 for the 3' region of Fhact1, 1000 and 1001 for the 5'region of Fhact4.4 and 611 and 612 for the 5'region of Fhact5.

#### *Marchantia polymorpha*

For *Marchantia* the identified actin genes were named Mpact1, Mpact4 and Mpact15. For all three sequences the 3'ends are lacking. So far for Mpact1 2229 bp, for Mpact4 3987 bp and for Mpact15 2174 bp of genomic sequences were isolated and sequenced. The lengths of the coding cDNA sequences isolated are 997 nt, 962 nt and 995 nt for Mpact1, Mpact4 and Mpact15 respectively. (Fig. 18 C). The sequence homologies within the *Marchantia* actin genes are a little bit lower than compared to the other two moss species, in a range between 78.3 and 85.5 % on the nucleotide level and between 94.7 and 96.1 % on the amino acid level (DNA STAR, MegAlign Program, Clustal V (weighted) sequence alignment). 5'upstream sequence of the ATG for all the three identified different *Marchantia* actin genes were isolated by iPCR and sequenced: 937 bp for Mpact1, 3025 bp for Mpact4 and 910 bp for Mpact15. The 5'regions of the the *Marchantia* actin gene homologous are amplified by PCR on genomic DNA from *Marchantia polymorpha* using the primer 950 and 951 for 5'Mpact1, 960 and 961 for Mpact4 and 970 and 971 for Mpact15. The intron-exon structure of the ORF was obtained by comparing the different actin gene sequences from the different moss species. The isolated 5' sequence of Mpact1 shows the consensus sequence for intron splice sites (aggt) at position -14 indicating the presence of a 5'intron as in the case of the other *Physcomitrella* and *Funaria* genes. Within the 5'upstream sequences of Mpact4 and Mpact15 no intron splice site consensus sequence is present, proposing the lack of 5'introns (fig. 18 C).

Comparison of of *P. patens*, *F. hygrometrica* and *M. polymorpha* actin genes

As mentioned above in general the homologies of nucleotide and protein sequences for the different isolated actin genes within one species is very high especially at the protein level. The homologies between the closely related moss species *Physcomitrella patens* and *Funaria hygrometrica* also appear to be very high. On the nucleotide level the actin genes show homologies between 86.9 and 96.3 % identity and on the amino acid level the range of homology is 95.5 to 99.7 %.

In contrast to that the more distant relation of the liverwort *Marchantia polymorpha* to the other both species is reflected in the lower homologies of the genes on the nucleotide level. The homologies between *Physcomitrella* and *Marchantia* actin genes is in the range of only 75.2 % and 78.8 % and between *Funaria* and *Marchantia* the homologies are in the range of 75.5 % to 80.4 %. On the amino acid level the homologies of the *Marchantia* actin genes vary between 93.0 % and 96.1 % compared to *Physcomitrella* and between 93.4 % and 96.7 % compared to *Funaria*.

#### Intron-exon structure (Fig. 18 A/B/C)

As indicated before the intron-exon structure of the *Physcomitrella* actin genes to a certain extent are similar to that of higher plants but also with clear differences. All isolated *Physcomitrella* actin genes contain a large 5' intron in the 5' untranslated region, which almost all of the investigated higher plants actins do. Only *Ppact1* contains 3 internal introns within the ORF reflecting the situation for example for all isolated actin genes from *Arabidopsis thaliana*. The ORF internal intron positions of *Ppact1* are also conserved compared to higher plant actin genes. On the contrary *Ppact3*, *Ppact5* and *Ppact7* contain only one internal intron within the ORF.

The same genomic structure can be found in the isolated *Funaria* actin genes with one extended 5' intron within the 5' UTR. *Fhact1* has the same conserved intron-exon structure as *Ppact1* whereas *Fhact4.4* and *Fhact5* contain only one internal intron within the ORF sequence. The isolated sequence of *Fhact5b* is too short to say something clear about the intron-exon structure but at least it does not contain the internal intron2 compared to *Fhact1* or

Ppact1.

In Marchantia the genomic structures of the isolated actin genes seem to be more different. It is important though, to indicate that the number of different actin genes in the three different moss species is not known and it could be that the three isolated actin genes from Marchantia do not represent the individual functionally homologous genes. It is likely that there are more than three actin genes present in Marchantia and more than four actin genes in Physcomitrella and Funaria.

However, the intron-exon structure of Mpact1 seems to be the same as in the case of Ppact1 and Fhact1 with a 5'intron within the 5'UTR and the conserved positions of the ORF internal introns 1 and 2. Mpact15 also contains the conserved ORF internal intron1 and intron2 but it does not have a conserved intron splice site at position -14 within the 5'UTR or at position -10 as found for the Physcomitrella or for some Arabidopsis actin 5'introns respectively, arguing for a lack of a 5'intron. The same situation is found for Mpact4, probably lacking a 5'intron. In addition Mpact4 also does not have the intron1 or the intron2 within the ORF, which is different from all isolated moss actin genes so far.

#### **Putative homologous moss actin genes**

Although the intron-exon structure of the different isolated actin genes from Physcomitrella and Funaria might propose conclusions about homologous genes between the two species one can not conclude this from the genomic structure. For example Ppact1 and Fhact1 share the same conserved intron-exon structure but it is not clear, as indicated before, whether there are more genes present in the genome of both plants which might have the same genomic structures. To give a statement on homologous genes also expression data would be required to propose functional homologies. Also from the sequence homologies of the proteins or the coding cDNA sequences it is not possible to make any assumptions about corresponding homologous genes between the species as they are too similar in general.

But in the case of *Physcomitrella* and *Funaria* it was interesting to find also very high sequence homologies within the non coding sequences regarding to the UTR sequences, intron sequences and promoter sequences. Therefore high homologies were found between Ppact1 and Fhact1 and between Ppact3 and Fhact5. In both cases the intron sequences showed unusual high conservation. In the case of Ppact1 and Fhact1 the homologies were as follows: 5'intron: 58 %; intron1: 64 %, intron2: 52 % and intron3: 55 %. In the case of Ppact3 and Fhact5 the homologies are for the 5'intron 51 % and intron1 shows 48 % identity.

For both cases also the isolated 5' promoter sequences show high homologies. Fig. 19 A shows a schematic comparison of the isolated promoter regions of Ppact1 and Fhact1. The transcription start is said to be at position 1, the first nt of the 5'promoter region is said to be -1. The isolated 267 bp of 5'promoter region of Fhact1 show an over all homology to the first 267 bp of the Ppact1 5'promoter region of 58 %. Within this sequence there are blocks of different homologies observable. The sequence between -267 and -129 shows a homology of 51 %. The following 29 bp show 62 % identity and within position -100 and -1 the homology is almost 70 %. Concerning these high sequence identities between the Ppact1 and Fhact1 intron and promoter sequences it is reasonable to put these two genes as the homologous genes in these two mosses. Another interesting aspect is the observation of the drop of expression observed between the different Ppact1:vegf deletion constructs (Fig. 15). The dramatic drop of expression appears to be between the -237 and the -82 deletion construct. This argues for an important function of the 5'promoter region between -129 and -1 as here the sequence of the promoter regions of Ppact1 and Fhact1 is highly conserved as just mentioned and the -82 deletion construct does not contain all of the highly conserved sequence but the -237 deletion construct does.

Highly conserved regions within the promoters of Ppact3 and Fhact5 can also be observed. In this case the promoter regions for both genes isolated are much longer. Therefore even more regions of homologies are found between the two 5'promoter regions (Fig. 19 B). In this case the promoter regions of Ppact3 from -1

to -2270 and of Fhact5 from -64 to -2325 show some interesting homology features. The difference in the TS position might be due to the fact that the 5'UTR of Fhact5 was determined experimentally and the one of Ppact3 was determined by analysing ESTs from database.

The sequence of Ppact3 between -2270 and -1876 shows only a 29 % low homology to the same sequence area of Fhact5 located between -2325 and -1948. Then an expanded region of about 1100 nt is following showing a very high homology of 82 %. The next 140 nt of Ppact3 and 152 nt of Fhact5 promoter show "only" 53 % homology. The sequence of Ppact3 located between -641 and -463 shows again high conservation of 76 % to the region between -705 and -528 of Fhact5. The following about 180 nt show again lower homology of 53 %. The last 288 bp of Ppact3 promoter sequence then are again more homolog with 73 % to the next 280 bp of Fhact5. These regions of different degrees of homologies between the two homologous genes might indicate the presence of regulative active elements within the 5'promoter region.

As for the case of Ppact1 and Fhact1 also here the expression analysis of the different Ppact3:vegf deletion constructs are interesting in this context (Fig. 17). Here a significant increase of the vegf expression level of the -2210, -995, -821 deletion constructs compared to the -523 deletion construct was observed. The three deletion construct which contain at least parts of the expanded homolog region between -1876 and -779 found in Ppact3 and Fhact5 reached levels about that of the 35 S promoter whereas the -523 deletion construct showed a 2 ½ fold increase of expression compared to the 35S promoter or the longer deletion constructs. This might argue for the presence of a negative regulator within this region of 82 % homology between Ppact3 and Fhact5.

In the case of Marchantia, no comparable sequence homologies could be found between the different actin genes from Physcomitrella and Funaria.

For the Fhact5 gene a construct containing 1157bp of the 5'untranscribed region fused to the hVEGF cDNA was made and used for

transient transformation experiments on *Physcomitrella* protoplasts. The amount of protein detected in this case was in the same range but slightly higher (up to 2 folds) as with the CaMV 35S promoter. The *Fhact5* gene presents the highest homology to the *PpAct3* gene, and interestingly both of the promoters showed a similar activity in *Physcomitrella* protoplasts during the transient assays.

## **2.5. Stable transgenic lines.**

The cassettes containing *Ppact1*, *Ppact5* and *Ppact7* 5' MEPRs driving the expression of the VEGF cDNA were introduced in the genome of *Physcomitrella* plants. For each of the MEPRs five to ten stably transformed plants were recovered and tested for the expression of rhVEGF. For these three MEPRs tested, expressed and secreted moss derived rhVEGF was detected in the supernatants of the cultures where the plants were growing (standard Knop medium), indicating that the MEPRs promote protein expression under non-inducing conditions (standard conditions) when they are integrated in other parts of the genome. The amount of protein that could be measured in those lines ranged from 7ngVEGF/mg moss dry weight until 53ngVEGF/mg moss dry weight, depending on the construct and the stable line.

One transgenic moss strain containing VEGF cDNA under control of *Ppact5* was used to perform bioreactor cultures. The amount of moss derived recombinant VEGF in the supernatant of bioreactor cultures measured by ELISA was 40-50ngVEGF/mg moss dry weight.

### **EXAMPLE 3: Cloning and analysis of *Physcomitrella patens* and *Funaria hygrometrica* ubiquitin genes and their expression promoting regions.**

Taking advantage of the presence of several EST sequences corresponding to polyubiquitin genes of *Physcomitrella*, specific oligos were designed to isolate the corresponding genomic sequences of the most abundantly present EST of the ubiquitin gene homologous sequence in the databases, named *Ppubq1*. 2146bp of 5' region of *Ppubq1* could be identified by iPCR. A 129bp transcribed 5' leader is present before the ORF starts, determined by



5' race. The 5' region of Ppubq1 is amplified by PCR on genomic DNA from *Physcomitrella patens* using the primers 777 and 602.

Vectors carrying different parts of promoter and 5'UTR region driving expression of the hVEGF cDNA, were constructed to analyse the activity of the the promoter during transient transformation of *Physcomitrella* protoplasts.

The results indicated a similar activity for this promoter to the Ppact5 promoter (or even higher). The constructs tested, 1.6Kb and 1.3Kb promoter fragments, reached expression levels around 4 times and almost 7 times higher than the CaMV 35S.

The ubiquitin gene from *Funaria*, Fhubq1, was identified by performing a 5' race PCR on *Funaria* total RNA with a primer derived from the Ppubq1 coding sequence. The isolated 5'UTR sequence and partial coding sequence was used to design primers for iPCR on genomic ligations of *Funaria hygrometrica*. This way 5' upstream sequence of the 5'UTR was identified. The 5' region is amplified by PCR on genomic DNA from *Funaria hygrometrica* using the primers 943 and 944.

**EXAMPLE 4: Cloning and analysis of *Physcomitrella patens* RBCS expression promoting regions.**

As putative candidates next to the actin, tubulin and ubiquitin genes the ribulose-1,5- biphosphate carboxylase/ oxygenase small subunit (rbcS) genes were taken into consideration. The different rbcS genes are encoded on the nuclear genome. The rbcS genes are members of a gene family. The rbcS genes are expressed basically in all green parts of plants able to fixate CO<sub>2</sub>. Therefore this gene family is of interest to get 5' and 3' flanking expression promoting regions of different rbcS genes from different mosses. As a first step *Physcomitrella* EST databases were analysed. It was found that the rbcS genes from *Physcomitrella patens* are organised in a gene family, consisting of 12 genes. The most abundantly present ESTs of the rbcS genes, named PprbcS12, was taken as a candidate to find it's 5' and 3' expression promoting sequences. Starting with the EST sequence data, 5' and 3' flanking regions of this gene was identified by iPCR and the

cloned 5' and 3' regions were sequenced. The 5' region is amplified by PCR on genomic DNA from *Physcomitrella patens* using the primers 839 and 858. The 3' region is amplified by PCR using the primers 904 and 901.

**In the enclosed Sequence Listing, the following sequences are given (Seq.ID.No/name of sequence/ 5' or 3' region relative to the protein encoding region):**

- 1 Pptub1 5'
- 2 Pptub1 3'
- 3 Pptub2 5'
- 4 Pptub2 3'
- 5 Pptub3 5'
- 6 Pptub3 3'
- 7 Pptub4 5'
- 8 Pptub4 3'
- 9 Ppact1 5'
- 10 Ppact1 3'
- 11 Ppact3 5'
- 12 Ppact3 3'
- 13 Ppact5 5'
- 14 Ppact5 3'
- 15 Ppact7 5'
- 16 Ppact7 3'
- 17 Fhact1 5'
- 18 Fhact1 3'
- 19 Fhact4.4 5'
- 20 Fhact5 5'
- 21 Mpact1 5'
- 22 Mpact4 5'
- 23 Mpact15 5'
- 24 Ppubq1 5'
- 25 Fhubq1 5'
- 26 PprbcS12 5'
- 27 PprbcS12 3'

Table 1: List of primers

No.	sequence(5'-3')	No.	sequence(5'-3')
35	ATCCAGGAGATGTTTCAGGCG	363	CATCTTGTCCAACCTACCGCGACCCGAACCC
36	CCGMACGCTGTCCATRGTYCC	364	AATCTCGAGTAGCATAAGATAAAGATGTTCTCTACC
38	ACATTGATGCGCTCCARCTGC	373	GGTAAAGCTTCTCGAGTGCAGTAGACGACAAAATG
40	GGBATGGACGAGATGGAGTTCAC	374	CATCTTGCTCAAGCTGTGCGAAGCTC
67	AGCACATGCACACCAATACGCTTGTGCGAATTC	395	ATCTCGAGGATCCATTCAACGGAGGATAAGT
69	GTCGTCATAGACGACAAGACCGGGATCCACAGC	408	CAACTCGAGATCGGTCTGTAAAGCCCTGTATTTG
70	TCAGTGTGTCCGTGAATCTCTCTCTCTGCTTTG	413	ATTTCTCGAGTTGTTGAATCATGTTAATTGCCAATGGT
71	CTGTGTTTCGGATTAGACTCCCCGTAGCCTTTGTG	511	TTACTCGAGACTCTACTAATTGACAAGTATG
89	TCGATTGGCGAGTTGCGAAGGAGGGCAAGG	547	GTCAAGATTGGAGGTTCCTTGAG
90	TGCCGTGCTCATCTTGAGTATGGCGTGTG	548	TCCATCTCGAGTACCTCCGCTGTGTGTTTCAAAG
91	CTGCAAGCAATGCGCACTGAAACAAGATGG	549	GTGCCTCGAGCCACATCCCGACCGCC
95	GACCTGGAAACCTGCACAATCAGCGCATAGA	550	AGCACCTCGAGTACTGCCCTAGTGCCCTAATC
113	TAGCATAAGATAAAGATGTTCTCTACC	602	CATCCTTACAGGAGTACTGG
149	CTCACCAGCCAATGGCTATGC	611	ATGCATGGCAAAACATCCCCTG
203	CCGTGGGACTTAGTTGTCTTCACTTC	612	CATGGAGATGAAATGTTCTG
204	GATCGAAATTTGCTGCTTGGCCTCCAC	777	TTAACTCGAGATACAAGAGTTATAAATCATATAC
205	TCGCAGGATGTGTCCTTAGTCGAGAA	839	ATATCTCGAGATGCATGTAAGATAATTCCAATTAGA
206	AACCTTCACGATTTCCACAAGCCACAC	858	CATTGCTAAAATCTCTCCACACTCGAATC
219	TTGATACTCGAGAAGTCCAAAATAATTTAATGATAC	901	ATATCTGCAGTCATGAAACTTTCATTATGTATC
223	CATCTTTCGCTAAGGATGATCTACAACGAG	904	ATATCGGCGCCGGAACGAAATTTGTGAGCTCTCT
225	CATCTTCAGTGTGCTCTACCTCACG	908	CTTTCGTGTTGCCCTCAAGAGTG
226	CTACTCGAGCACATATAAATACTGCCCTAGTGCC	909	CATTTCCTTAATACGGACCTGCC
291	GACAGATCTCCTTAGTCGAGAAGCGCGGGACGTG	943	ATATCTCGAGGAATTCATTTCCATTAAACGAGATATGAC
292	GACCCGTGGGACTTAGTTGTCTTCACTTC	944	CATCTTCACAACGCTTTATCACTTC
296	GCTGCTCTTCTCGTGATTTGTCT	950	CATATGCGTACGGAGTTGTGG
297	CATTCCACCCCTTCTCTCTCTC	951	TTTCGCGAAGTTACCTAACC
298	GTTTTGTGGCTCTTCTCTTG	960	TCATGATGTTAAGCGTTTTC
299	ATCGCTTCTCGACTCTTCTTCC	961	GTTAACGAAGGAGGTGTCCG
300	GTTACGCTCGCAATGCGTACT	970	AAGCTTAGCAAGCAGCTCTCGCAG
320	AACCTTCTGCTGTCTTGGGTGCATTG	971	ATCGACGATAGACTGCAAGCC
321	GACCTGCAGGCATCGAGCTTGTAAATCATGGTCATAG	983	AGGAGTGTACACATCTTTTAC
332	CATTTCCTTAATACCGACCTGCCCAACCA	984	GGCTAAGACGACGCAATTCGTG
333	CATGGAGAAGAAATACTCTGCACATCAAAAG	1000	GGATCCGAGAGGAAAGAGAGAG
334	CATTATTTAATACGGACCTGCACAACAAC	1001	CGCTTACAATGATCTGTCATAG
335	CATTTTTTAGAATGATCTACAGGAGTTC	10R	TCGTTGAATCCATCTCTGTCAT
336	AGATCTGGCAAGTTCCCTTCG	8F	CGGTACCTACAAGGGCTCTCG
337	GAAGAGAAGGAAGGGTGGGAATG	9F	TGGGACGTATCAGGGTACGTCT
338	GGAAGAAGAGTCGAGAAGCGAT	F7	TATCCGGAGGTTCCCGCAGACC

REFERENCES:

- Altschul et al., Nucleic Acids Research 25 (1997), 3389-3402.
- Berlin et al., Genetic modification of plant secondary metabolism: Alteration of product levels by overexpression of amino acid decarboxylases, in: Advances in Plant Biology, Studies in Plant Science, Vol. 4, pp 57-81, DdY Ryu and S Furasaki (eds), Elsevier, Amsterdam 1994.
- Cove et al. Trends Plant Sci. 2 (1997), 99-105
- Engel, Am. J. Bot. 55 (1968), 438-446
- EP 1 206 561 A
- Frahm "Biologie der Moose" (2001), Spektrum Akad. Verlag
- Gorr (1999) Biotechnologische Nutzung von *Physcomitrella patens* (Hedw.) B.S.G.. Dissertation, Universität Hamburg.
- Grimsley et al., Mol. Gen. Genet. 154 (1977), 97-100
- Hiwatashi et al., Plant J. 28(1) (2001), 105-116
- Hohe et al., Plant Sci. 163 (2002), 69-74
- Holtdorf et al., Plant Cell Rep. 21 (2002), 341-346
- Liebich et al., Nucleic Acids Research 30 (2002), 3433-3442.
- McElroy et al., Mol. Gen. Genet. 231 (1991), 150-160
- Reski et al., Planta 165 (1985), 354-358
- Reski et al., Mol. Gen. Genet. 244 (1994), 352-359
- Reski, Planta 208 (1999), 301-309
- Reutter et al., Plant Tissue Culture and Biotechnology 2 (1996), 142-147
- Richter et al., Gene 290 (2002), 95-105
- Rothnie et al., Plant Mol Biol. 32 (1996), 43-61.
- Rother et al., J. Plant Physiol. 143 (1994), 72-77
- Sambrook et al. (1989), Molecular Cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press
- Schaefer et al., Mol Gen Genet 226 (1991), 418-424
- Schaefer (2001) Principles and protocols for the moss *Physcomitrella patens*. <http://www.unil.ch/lpc/docs/PPprotocols2001.pdf>
- Schaefer, Annu. Rev. Plant Biol. 53 (2002), 477-501
- Schlink et al., Plant Mol. Biol. Rep. 20 (2002), 423a-423f
- Strepp et al., Proc. Natl. Acad. Sci. USA 95 (1998), 4368-4373
- Töpfer et al., NAR 15 (1987), 5890
- Watson et al., "Recombinant DNA" (1992), Chapter I.1 and 2
- Zeidler et al., J. Plant Physiol. 154 (1999), 641-650

Claims:

1. Isolated nucleic acid molecule encoding a moss expression promoting region (MEPR).
2. Isolated nucleic acid molecule according to claim 1, characterised in that the MEPR is selected from MEPRs of *Physcomitrella*, *Funaria*, *Sphagnum*, *Ceratodon*, *Marchantia* and *Sphaerocarpos*, preferably of *Physcomitrella patens*, *Funaria hygrometrica* and *Marchantia polymorpha*.
3. Isolated nucleic acid molecule according to claim 1 or 2, characterised in that the MEPR is selected from the Seq. ID Nos. 1 to 27 or expression promoting fragments thereof.
4. Isolated nucleic acid molecules according to any one of claims 1 to 3, characterised in that it comprises a moss promoter and preferably a 5'-UTR region and/or a 5'-intron and/or a 3'-UTR
5. Isolated nucleic acid molecules according to any one of claims 1 to 4, characterised in that the MEPR has an expression promoting activity being at least equal to the expression promoting activity of cauliflower mosaic virus (CaMV) 35S promoter.
6. Isolated nucleic acid molecules according to any one of claims 1 to 5, characterised in that the MEPR has an expression promoting activity being at least 200 %, preferably being at least 500 %, especially being at least 1000 %, of the expression promoting activity of cauliflower mosaic virus (CaMV) 35S promoter.
7. Isolated nucleic acid molecules according to any one of claims 1 to 6, characterised in that it further comprises a coding region for a recombinant polypeptide product, said coding region being under the control of the MEPR.
8. Isolated nucleic acid molecules according to any one of claims 1 to 7, characterised in that it further comprises a selection marker.

9. Isolated nucleic acid molecules according to any one of claims 1 to 7, characterised in that it further comprises sequences which are homologous to genomic sequences of the species to be transformed thereby allowing targeted integration in this species.

10. Isolated nucleic acid molecules according to any one of claims 1 to 5, characterised in that is provided as an antisense or ribozyme molecule

11. Process for the expression of a recombinant polypeptide product in an eukaryotic host cell comprising the following steps:

- providing a recombinant DNA cloning vehicle comprising an isolated nucleic acid molecule encoding an MEPR according to any one of claims 1 to 10 and optionally a coding region for said recombinant polypeptide product, said coding sequence being under the control of the MEPR of said nucleic acid molecule in said host,
- transforming said eukaryotic host cell which does not naturally harbour said coding sequence in a way that it is under the control of said MEPR,
- culturing the transformed eukaryotic host cell in a suitable culture medium,
- allowing expression of said recombinant polypeptide and
- isolating the expressed recombinant polypeptide.

12. Method according to claim 11, characterised in that said eukaryotic host cell is selected from plant cells, preferably moss cells, especially *Physcomitrella patens* cells.

13. Method according to claim 11 or 12, characterised in that said host cell is a protonema moss tissue cell.

14. Method according to any one of claims 11 to 13, characterised in that the culture medium is free from added phytohormones thereof.

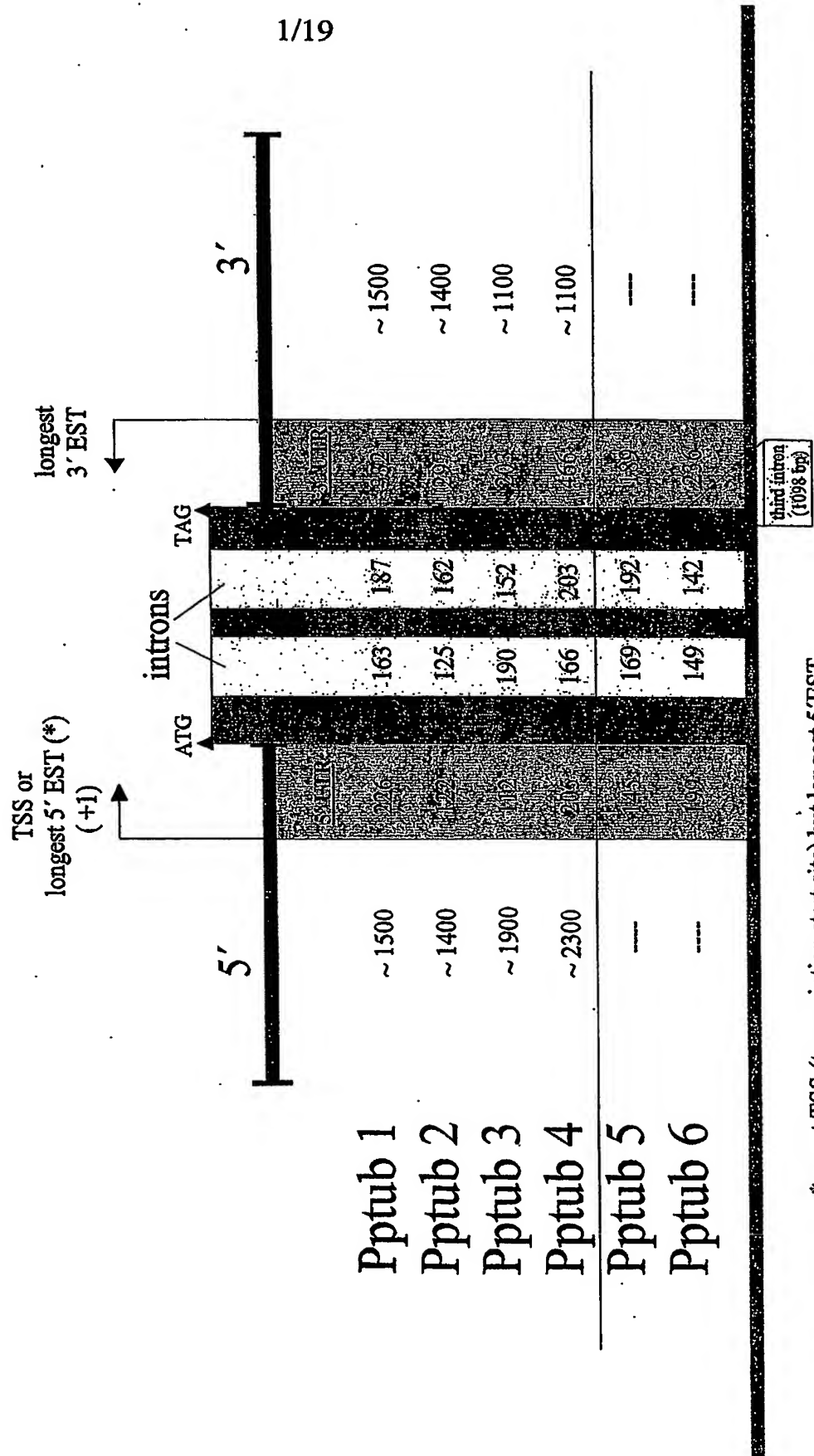
15. Method according to any one of claims 11 to 14, characterised in that the cell is selected from moss cells of the group Physcomitrella, Funaria, Sphagnum, Ceratodon, Marchantia and Sphaerocarpos.
16. Method according to any one of claims 11 to 15, characterised in that the host cell expresses said recombinant polypeptide product transiently.
17. Use of an isolated nucleic acid molecule encoding an MEPR according to any one of claims 1 to 10 for industrially producing a polypeptide, especially for providing recombinant cells producing said polypeptide.
18. Use of an isolated nucleic acid molecule encoding an MEPR according to any one of claims 1 to 10 for expression of a moss polypeptide, the expression of said moss polypeptide being not naturally controlled by said MEPR, especially for providing recombinant moss cells expressing said polypeptide.
19. Use of an isolated nucleic acid molecule encoding an MEPR according to any one of claims 1 to 10 for screening and defining consensus sequences for expression promoting regions.
20. Use of an isolated nucleic acid molecule encoding an MEPR according to any one of claims 1 to 10 for recombinant expression of postrtranslationally modifying proteins, especially for the production of posttranslationally modified proteins.
21. Use of an isolated nucleic acid molecule encoding an MEPR according to any one of claims 1 to 10 for in vitro expression of recombinant proteins.
22. Use of an isolated nucleic acid molecule encoding an MEPR according to any one of claims 1 to 10 for recombinant expression of metabolism modifying proteins.

Summary:

Moss Expression Promoting Regions

Disclosed are isolated nucleic acid molecules encoding moss expression promoting regions (MEPRs) as well as a method for producing recombinant polypeptides using such MEPRs.



Fig. 1:  $\beta$ -tubulin genes in *Physcomitrella patens*

\* = not TSS (transcription start site) but longest 5' EST  
all numbers in bp

Fig. 2: Analysis of expression promoting regions of  $\beta$ -tubulins in *Physcomitrella patens*

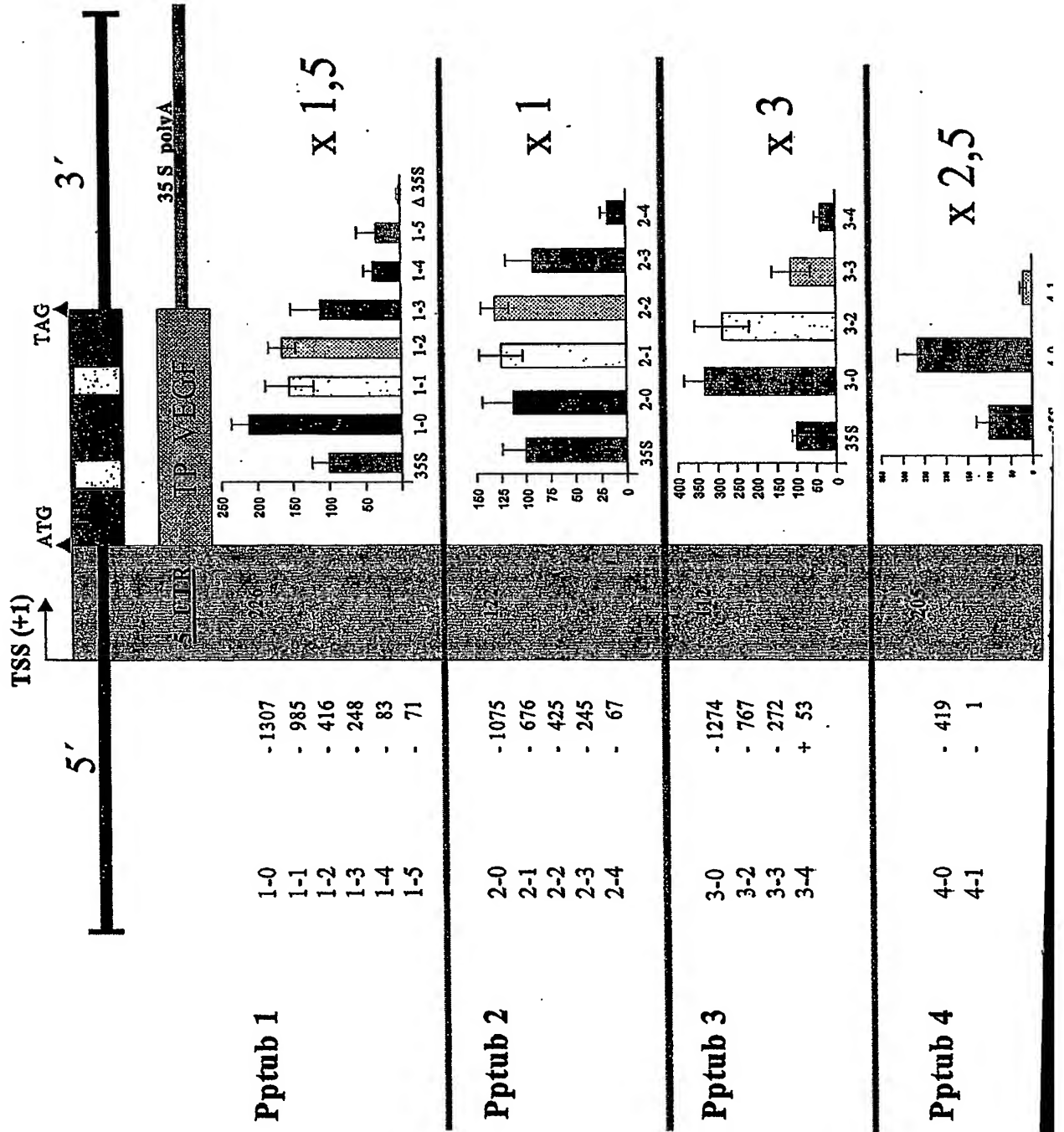
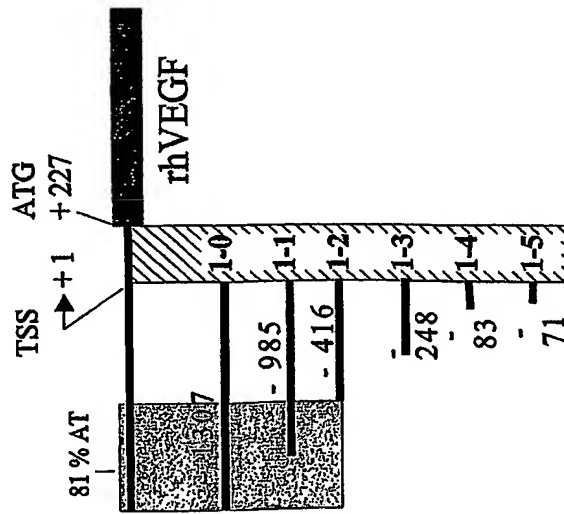


Fig. 3: Analysis of expression promoting regions of Pptub 1 by transient transformation of rhVEGF constructs



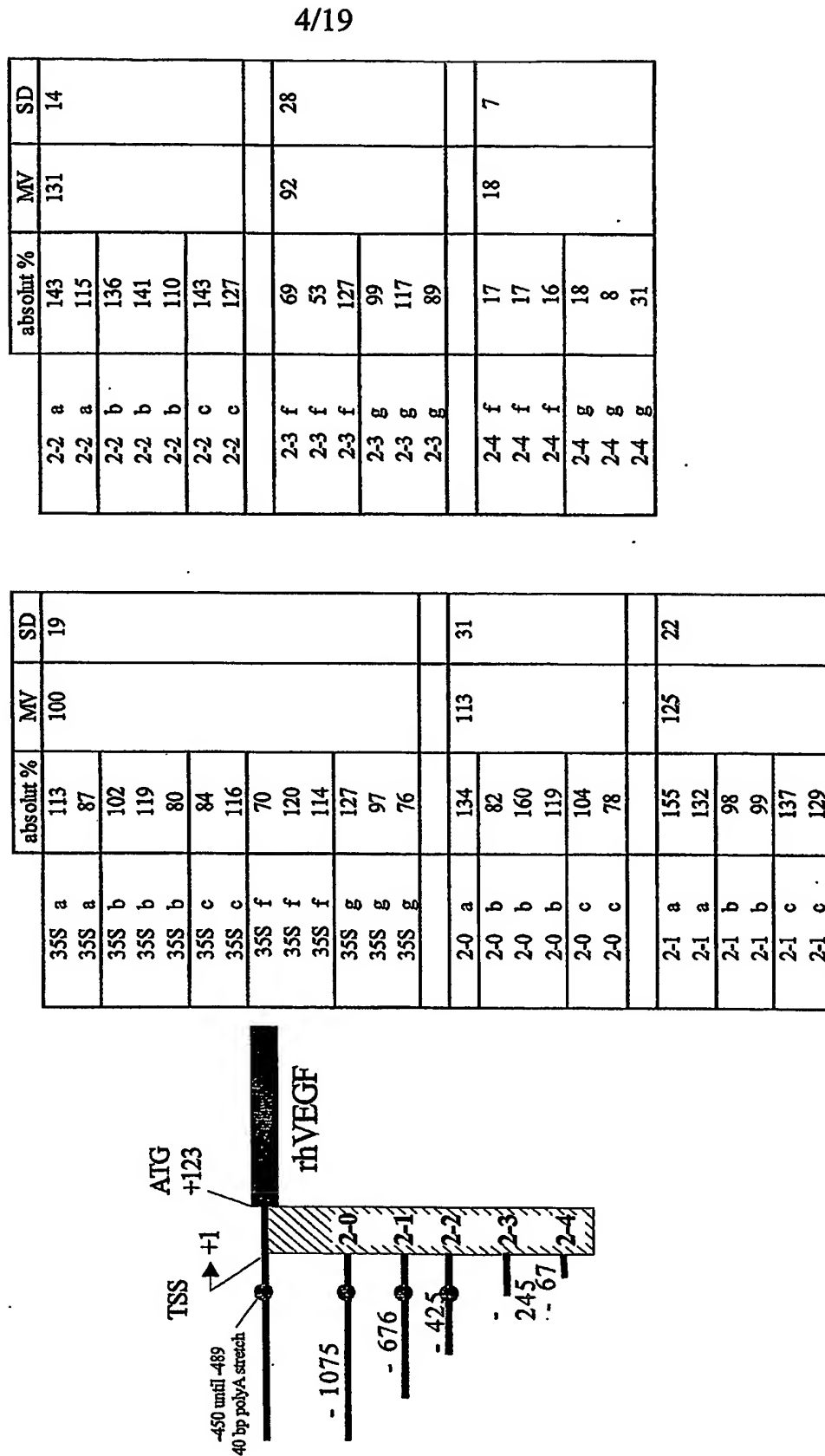
	absolut %	MV	SD
35S a	122	113	54
35S a	80		
35S a	98		
35S b	98		
35S b	309		
35S b	119		
35S c	107		
35S c	89		
35S c	104		
35S d	133		
35S d	99		
35S d	69		
35S e	93		
35S e	68		
35S e	139		
35S f	147		
35S f	78		
35S f	75		
1-0 a	247	212	23
1-0 a	197		
1-0 a	195		
1-0 c	234		
1-0 c	195		
1-0 c	204		
1-1 a	145	155	34
1-1 a	129		
1-1 b	165		
1-1 b	135		
1-1 b	144		
1-1 d	119		
1-1 d	179		

	absolut %	MV	SD
1-2 a	184	165	19
1-2 a	150		
1-2 a	173		
1-2 c	148		
1-2 c	187		
1-2 c	148		
1-3 d	150	111	41
1-3 d	85		
1-3 d	56		
1-3 e	119		
1-3 e	94		
1-3 e	164		
1-4 c	44	39	12
1-4 c	46		
1-4 c	53		
1-4 f	40		
1-4 f	33		
1-4 f	18		
1-5 f	6	34	26
1-5 f	9		
1-5 f	25		
1-5 e	36		
1-5 e	59		
1-5 e	70		
delta 35S e	5	2	3
delta 35S e	0		
delta 35S e	0		

a - f: different transformations  
SD = standard deviation  
35S mean value (MV)  
of each transformation was set to 100%

Fig. 4: Analysis of expression promoting regions of Pptub 2 by transient transformation of rhVEGF constructs

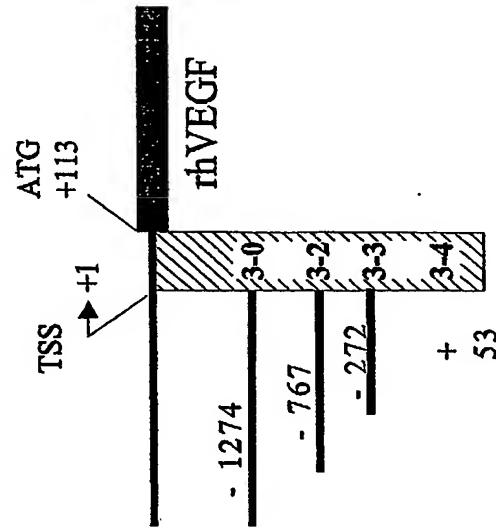


a - b, f and g: different transformations

SD = standard deviation

35S mean value (MV) of each transformation was set to 100%

Fig. 5: Analysis of expression promoting regions of Pptub 3 by transient transformation of rhVEGF constructs



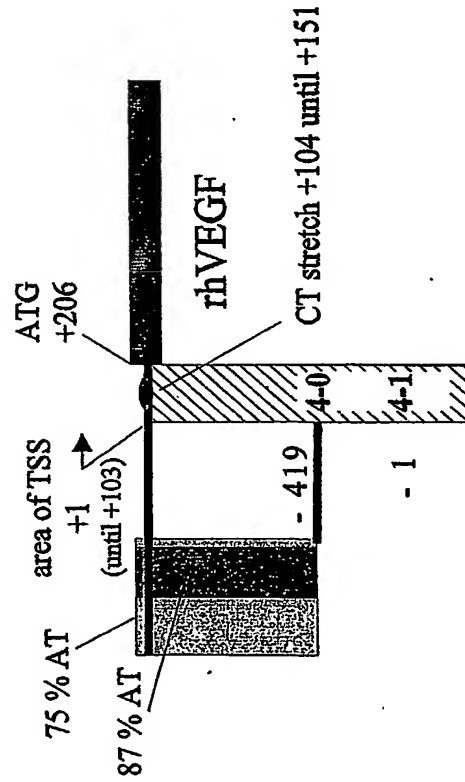
a - b, d and e: different transformations

SD = standard deviation

35S mean value (MV) of each transformation was set to 100%

	absolute %	MV	SD
35S a	100	100	11
35S a	95		
35S a	105		
35S b	118		
35S b	81		
35S b	101		
35S d	94		
35S d	112		
35S d	94		
35S e	93		
35S e	89		
35S e	118		
3-0 e	293	332	52
3-0 e	251		
3-0 e	353		
3-0 b	387		
3-0 b	330		
3-0 b	379		
3-2 a	231	287	69
3-2 a	239		
3-2 a	247		
3-2 b	399		
3-2 b	348		
3-2 b	259		
3-3 b	138	112	49
3-3 b	104		
3-3 b	191		
3-3 d	44		
3-3 d	96		
3-3 d	101		
3-4 a	27	37	15
3-4 a	16		
3-4 a	46		
3-4 b	55		
3-4 b	50		

Fig. 6: Analysis of expression promoting regions of Pptub 4  
by transient transformation of rhVEGF constructs

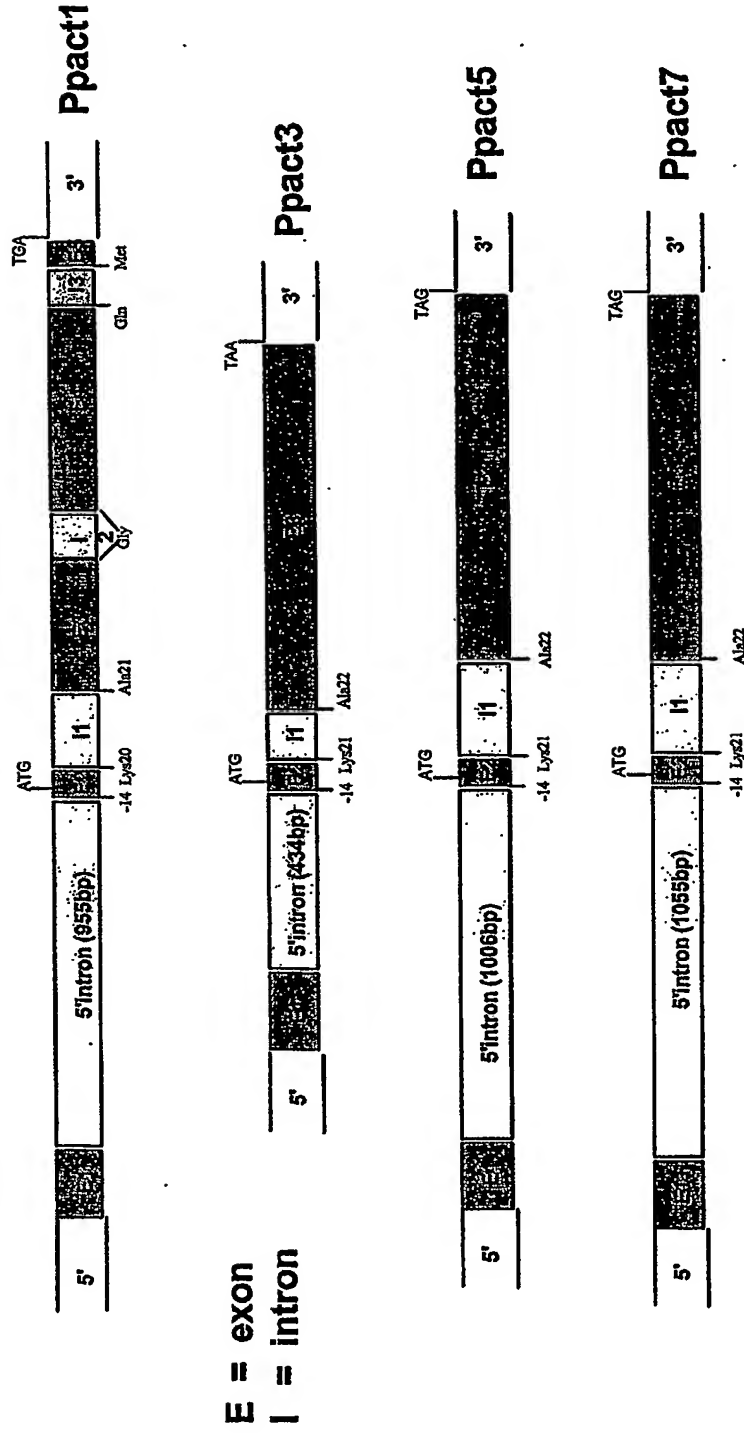


a and c: different transformations

SD = standard deviation

35S mean value (MV) of each transformation was set to 100%

	absolut %	MV	SD
35S a	63	100	30
35S a	95		
35S a	141		
35S c	70		
35S c	121		
35S c	109		
4-0 a	290	265	45
4-0 a	322		
4-0 a	229		
4-0 c	210		
4-0 c	273		
4-1 a	25	20	8
4-1 a	22		
4-1 a	5		
4-1 c	19		
4-1 c	30		
4-1 c	18		



### 5'sequences resulting from iPCR:

- Ppact1: 2973 bp until ATG: 1824 bp promoter / 955 bp 5' intron
- Ppact3: 3091 bp until ATG: 2270 bp promoter / 434 bp 5' intron
- Ppact5: 3095 bp until ATG: 1909 bp promoter / 1006 bp 5' intron
- Ppact7: 3069 bp until ATG: 1805 bp promoter / 1055 bp 5' intron

**Fig. 7: Genomic structure of *Physcomitrella patens* actin genes.**

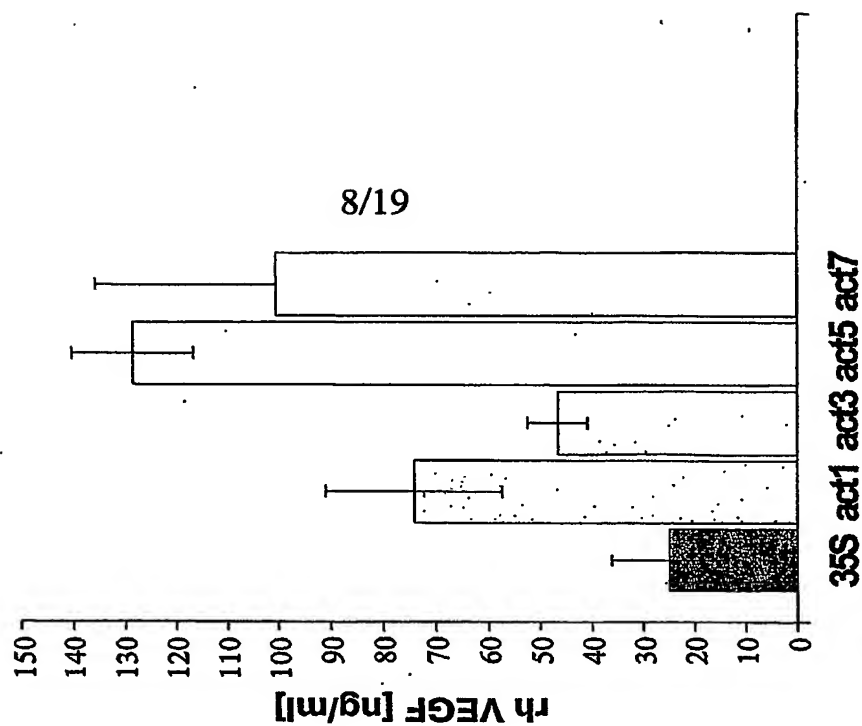
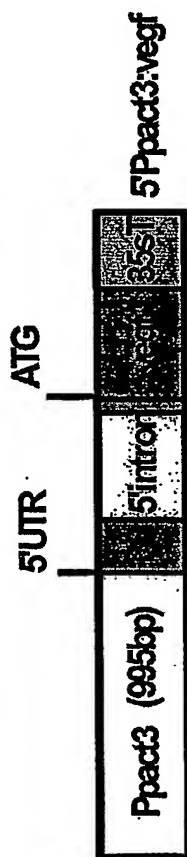
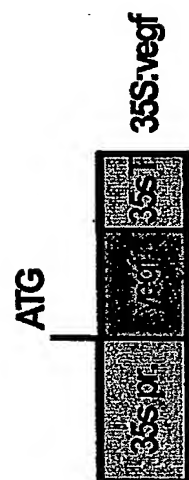


Fig. 8: Comparison of the expression activity of the different 5' actin regions.



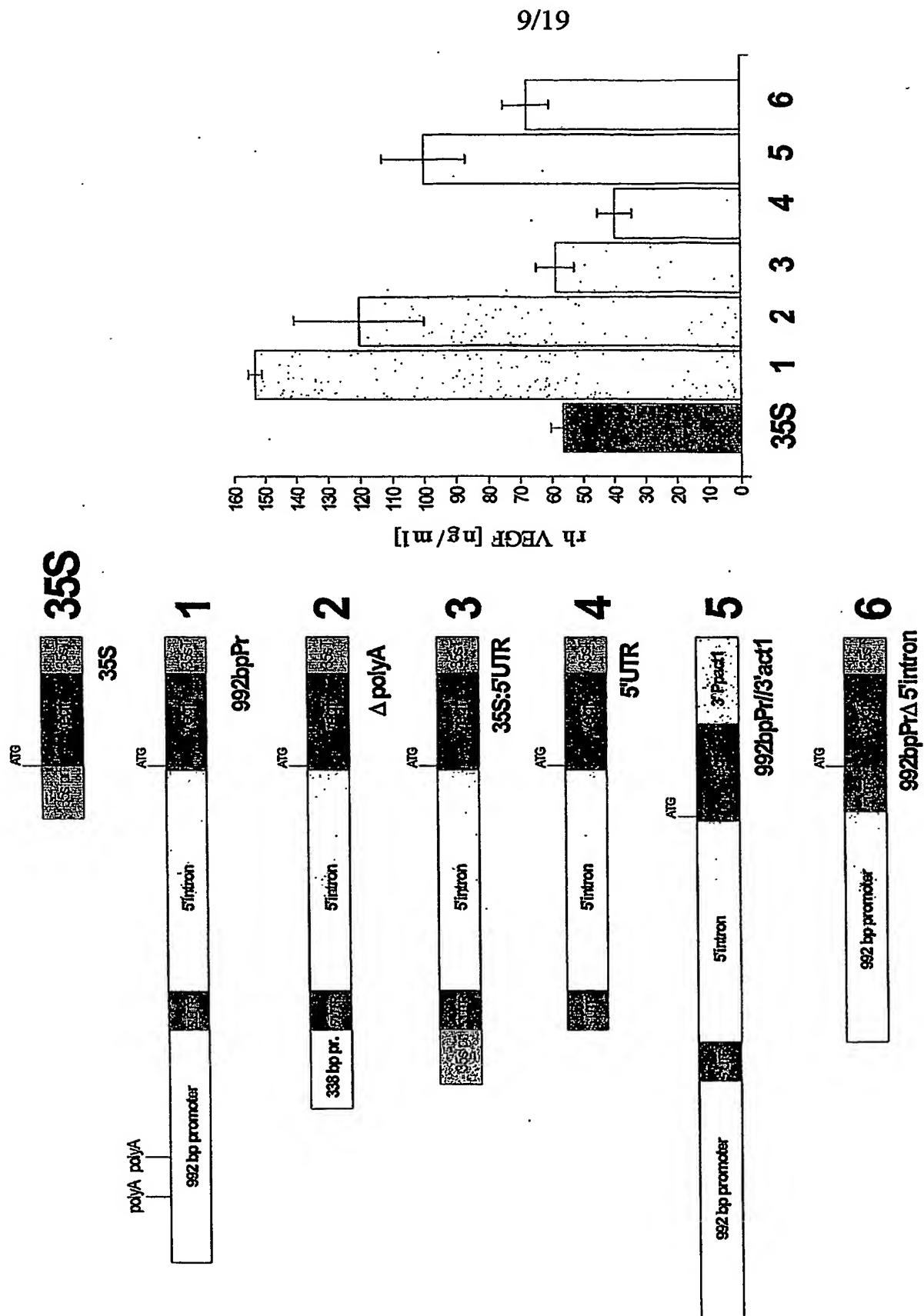
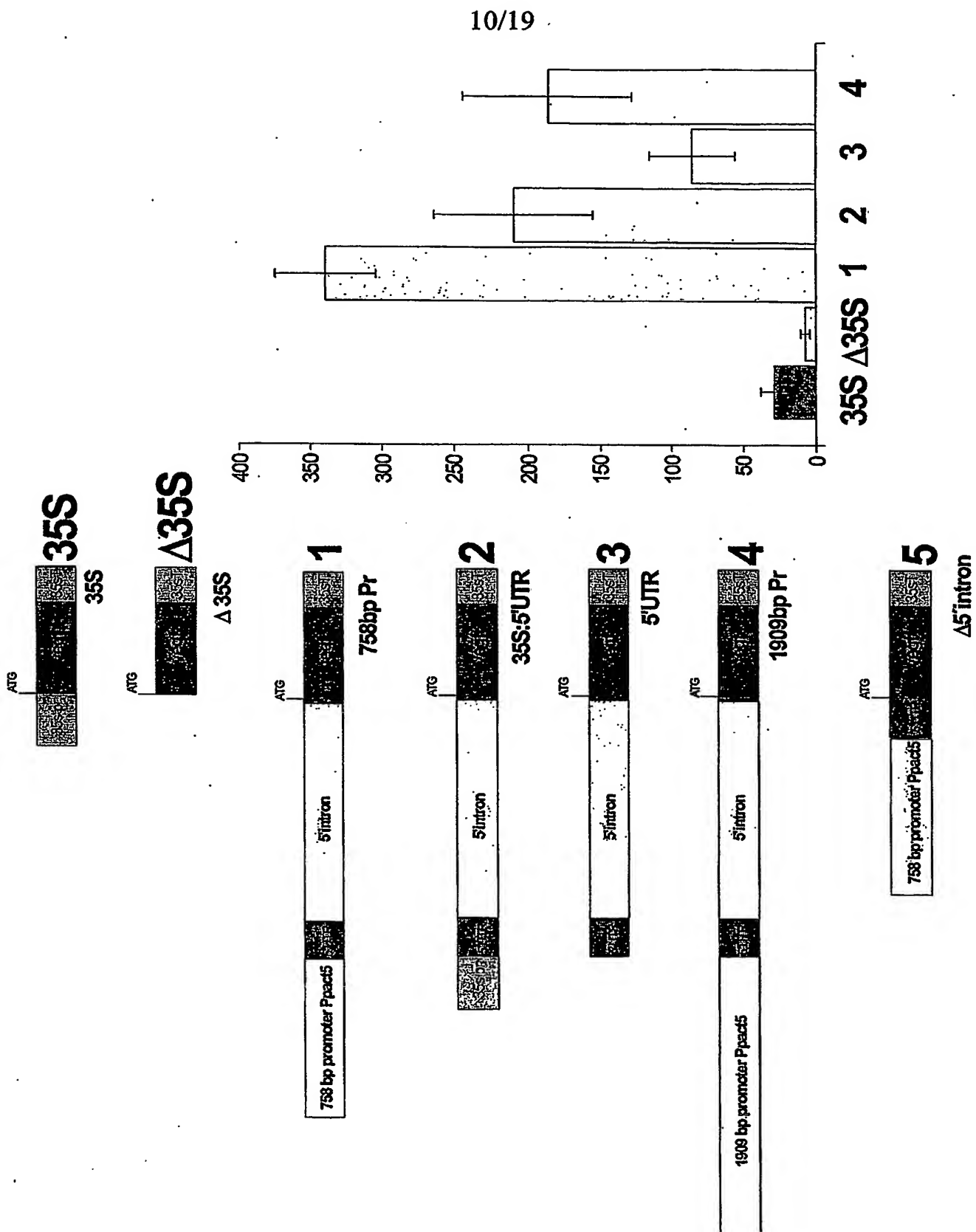


Fig. 9: Ppact1 constructs.



**Fig. 10: Ppact 5 constructs.**

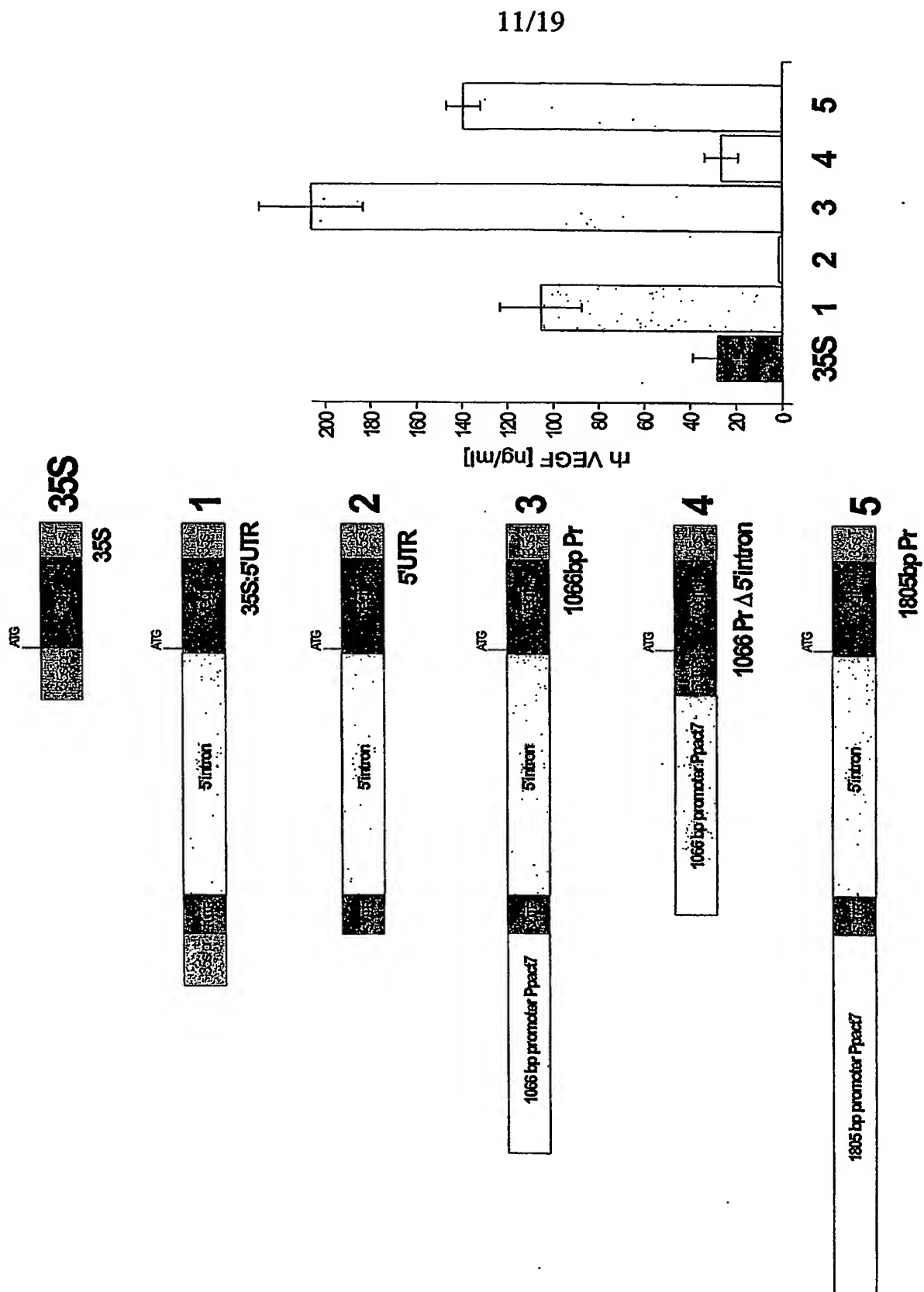
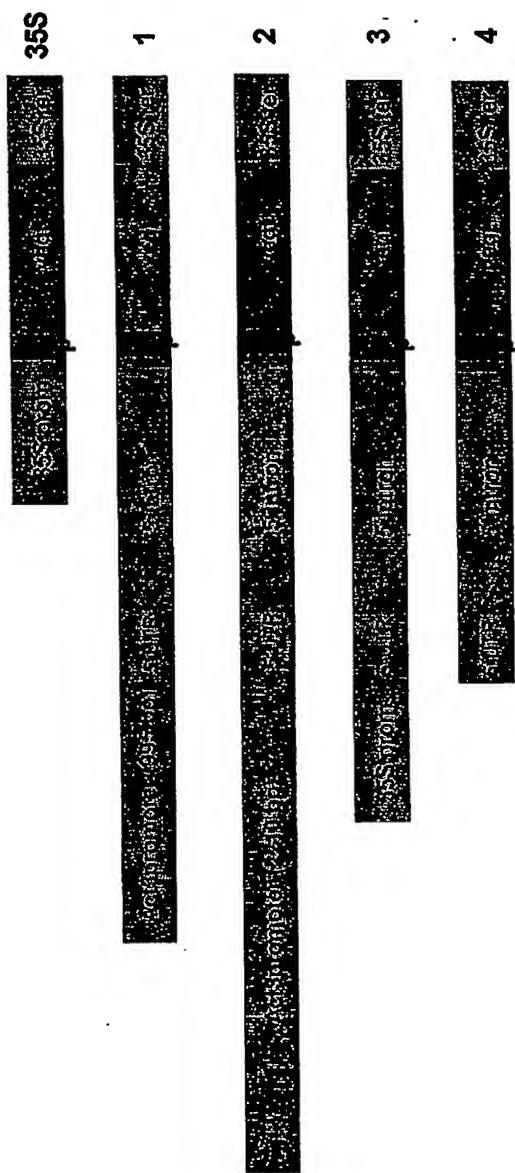


Fig. 11: Ppact 7 constructs.



12/19

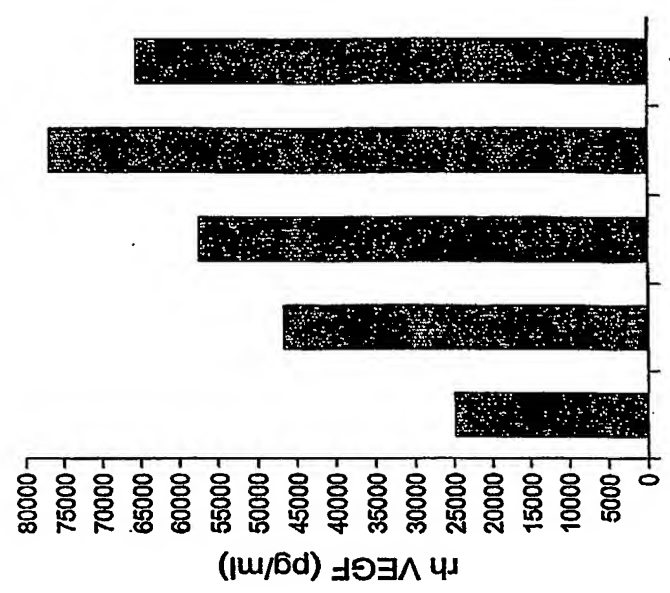


Fig. 12: Pp act3::vegf constructs.

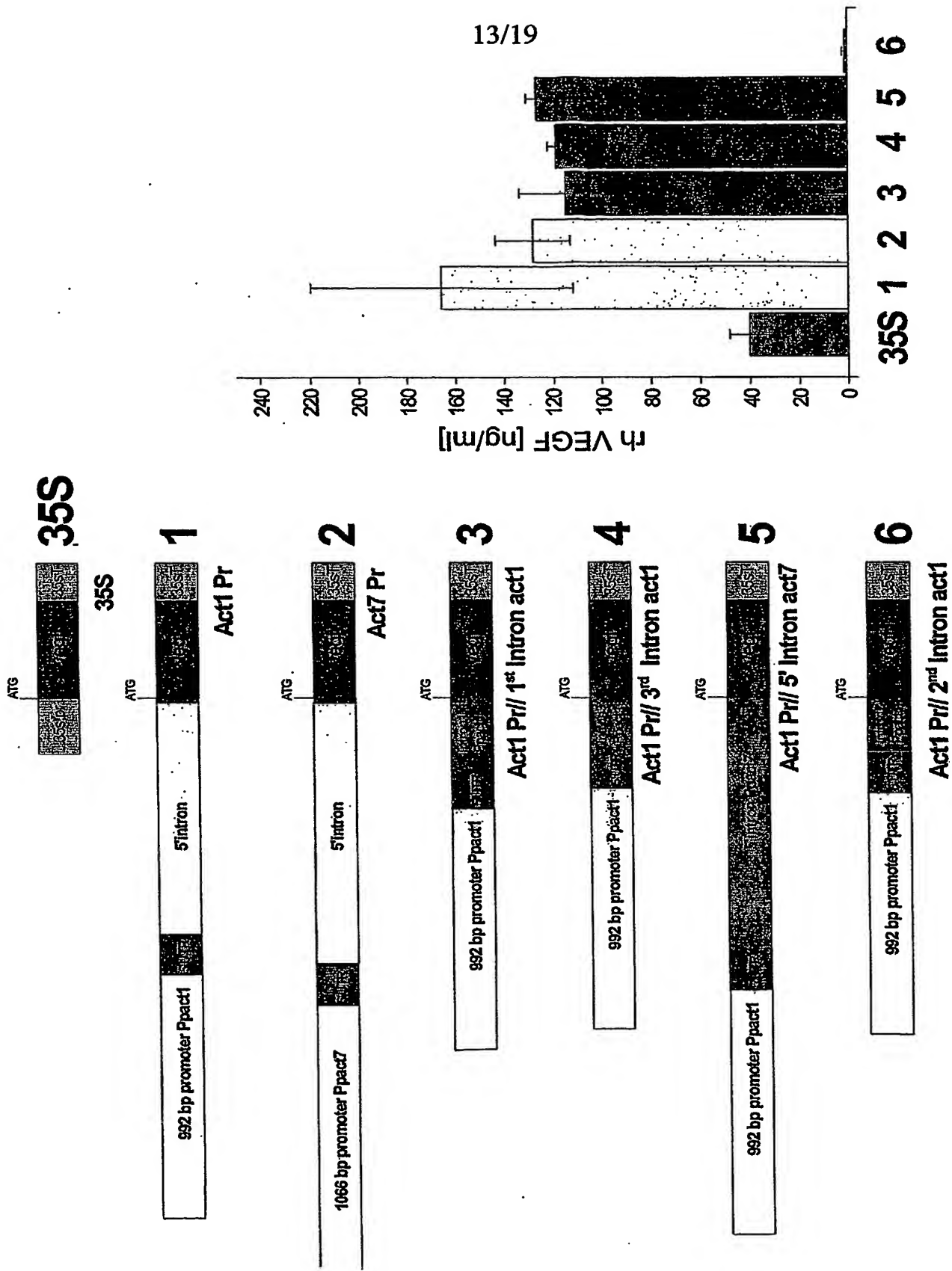
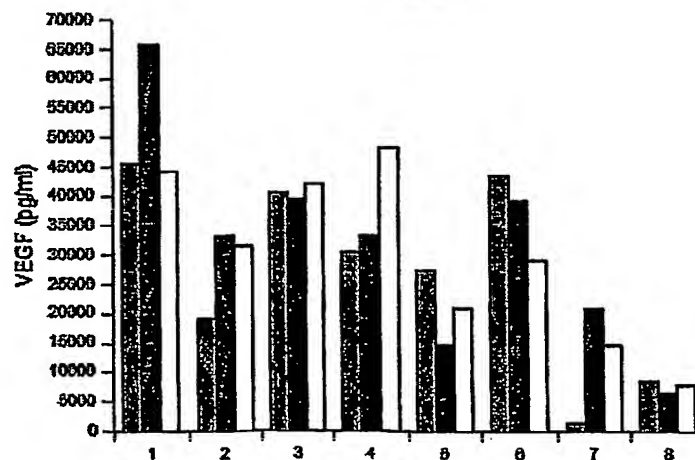
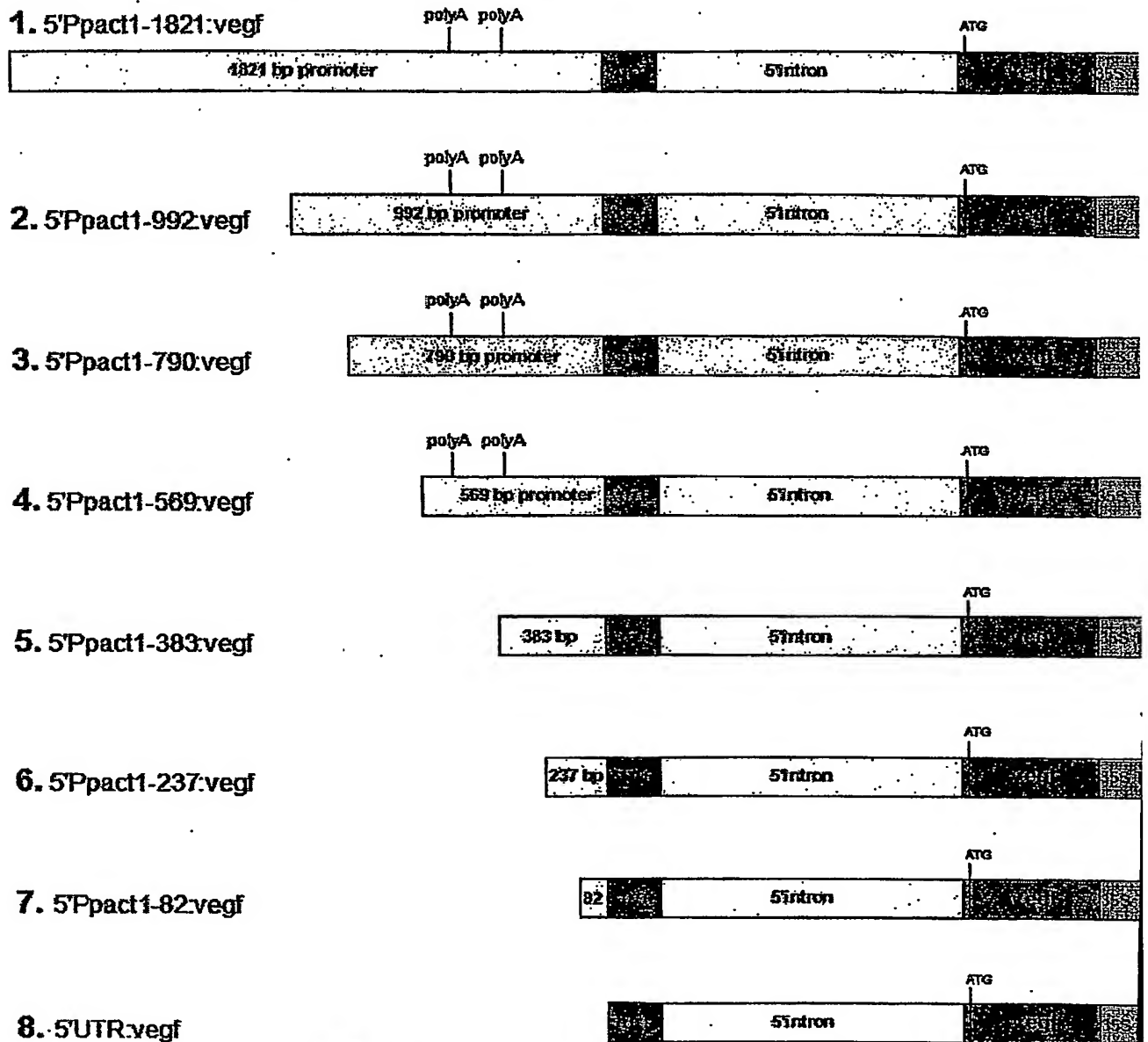
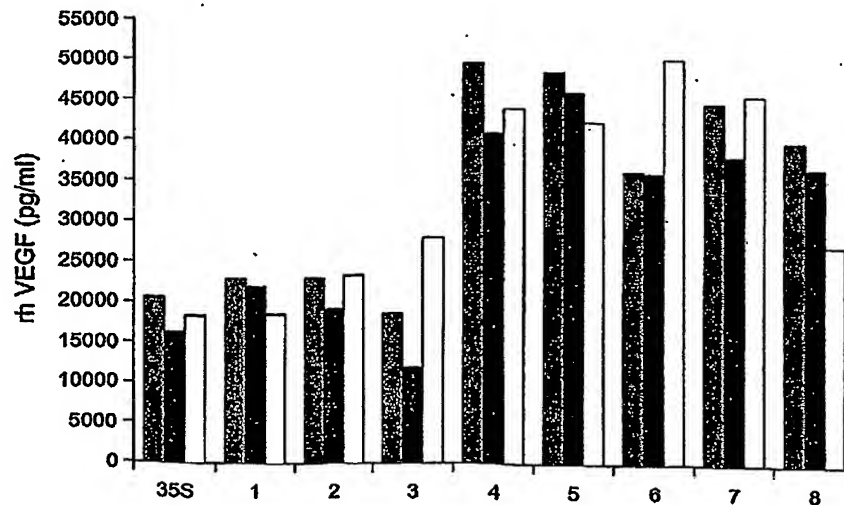
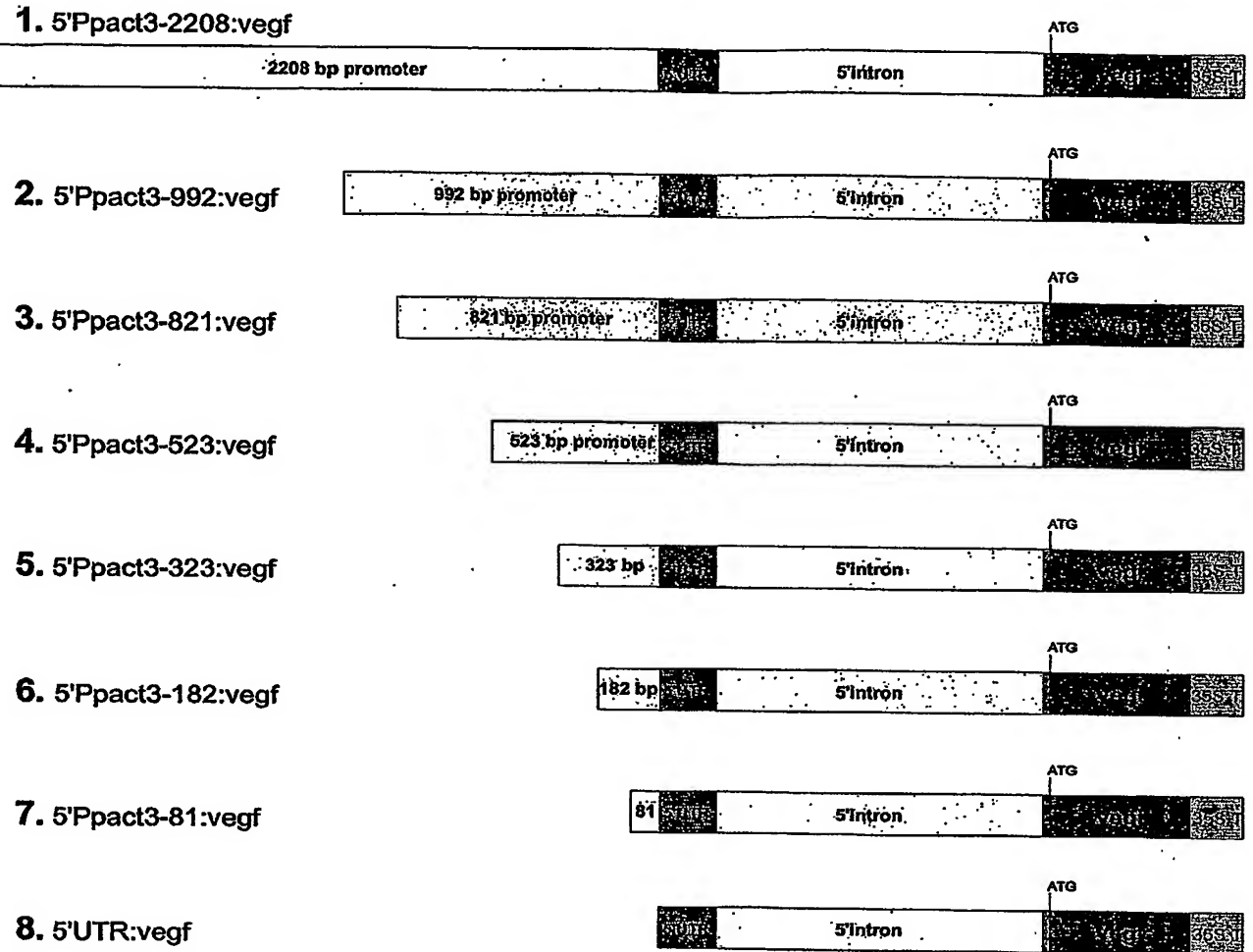
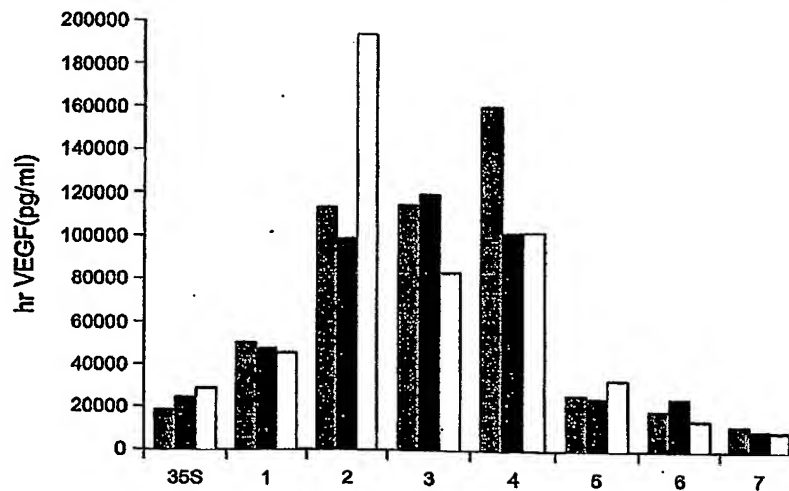
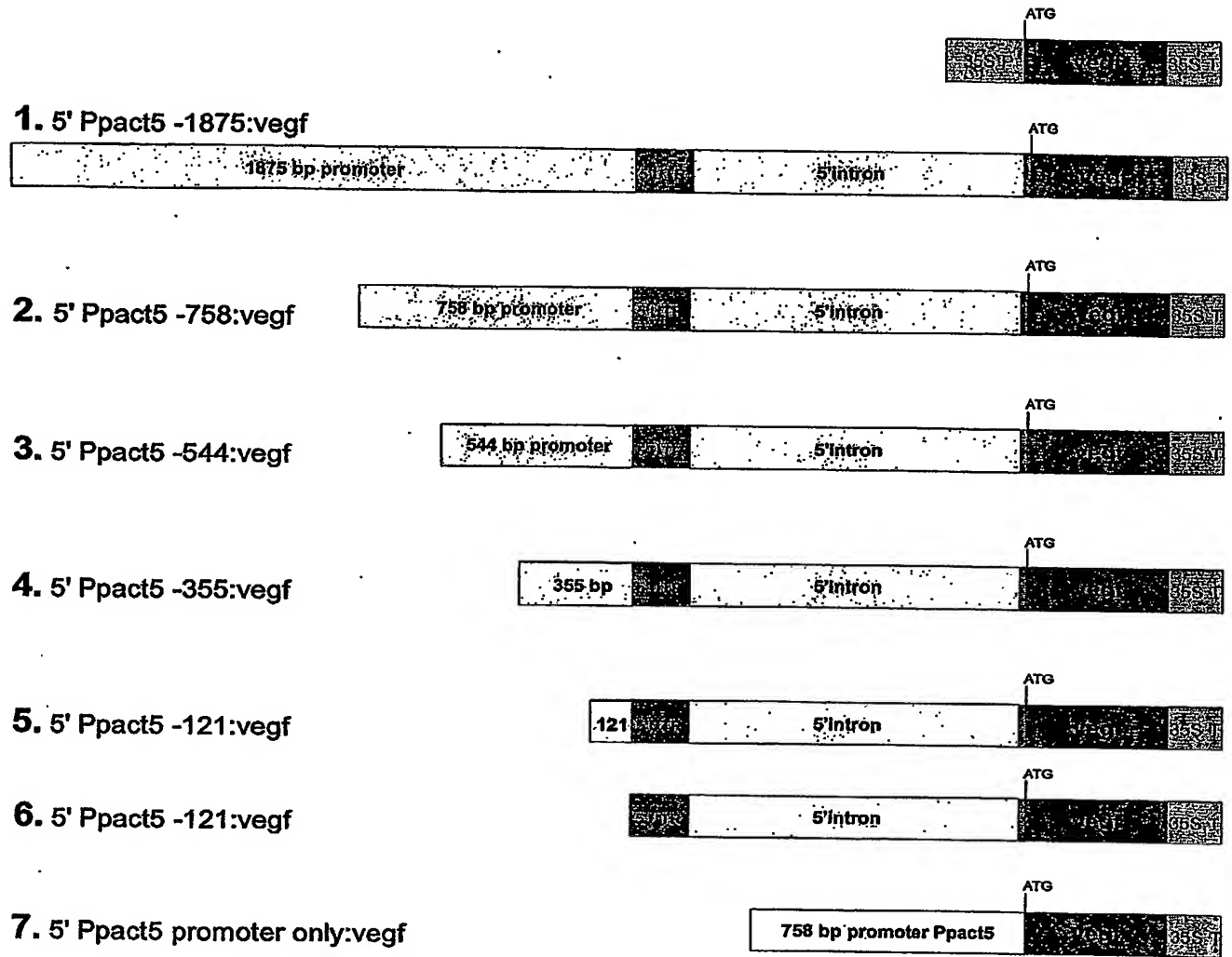


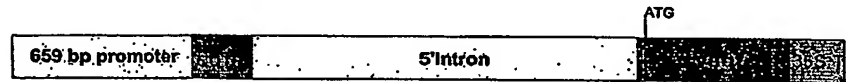
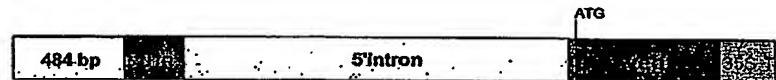
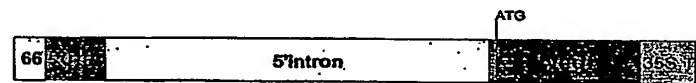
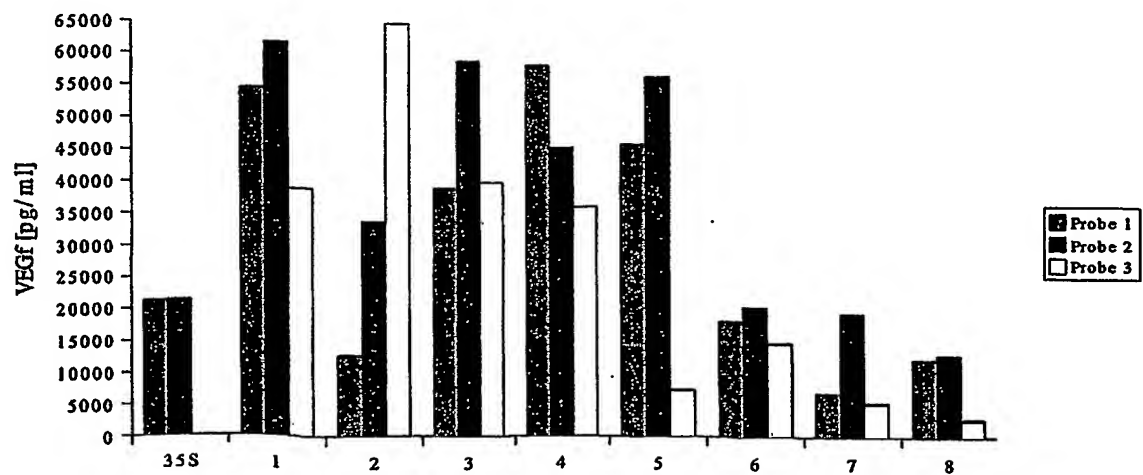
Fig. 13: Ppact1 promoter:5' intron substitutions.

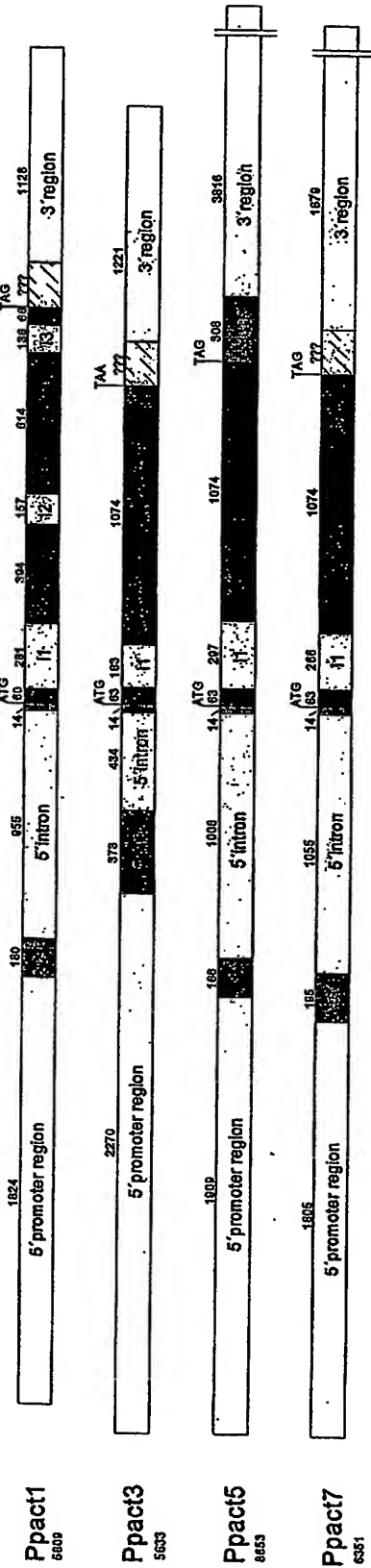
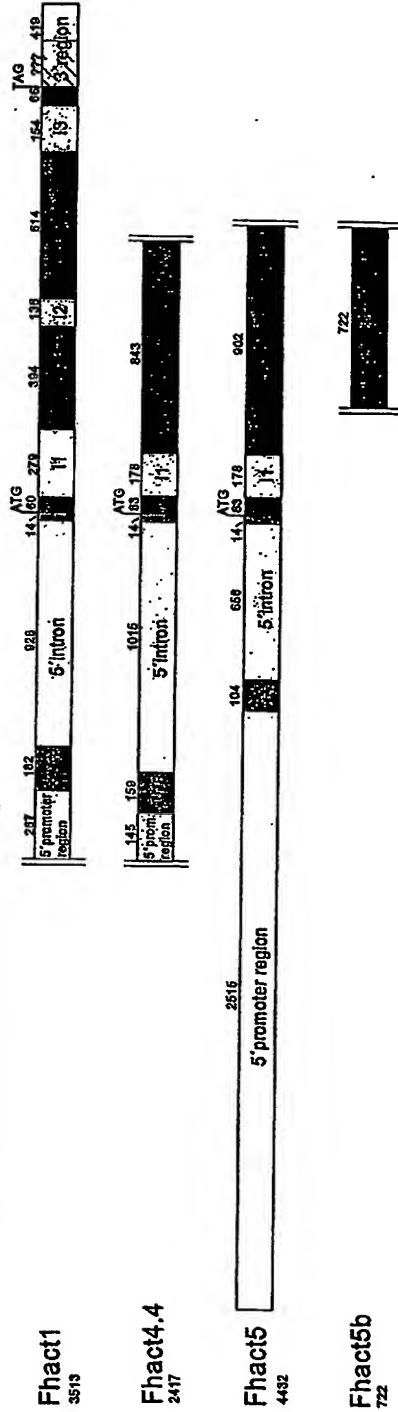
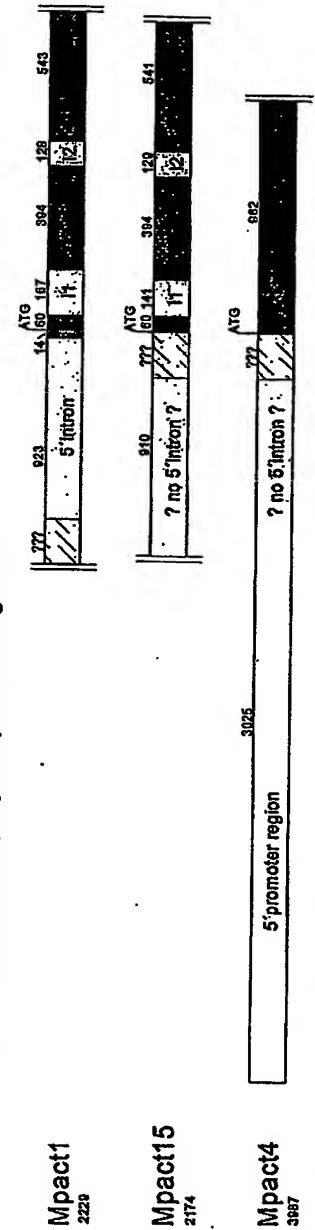
**Fig. 14: Ppact1 promoter:vegf deletion constructs.**

**Fig. 15: Ppact3 promoter:vegf deletion constructs.**

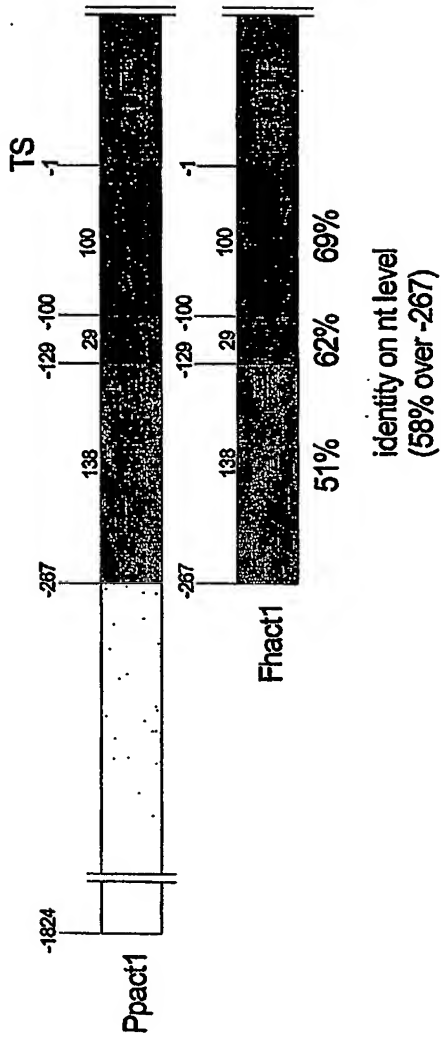
**Fig. 16: Ppact5 promoter:vegf deletion constructs.**



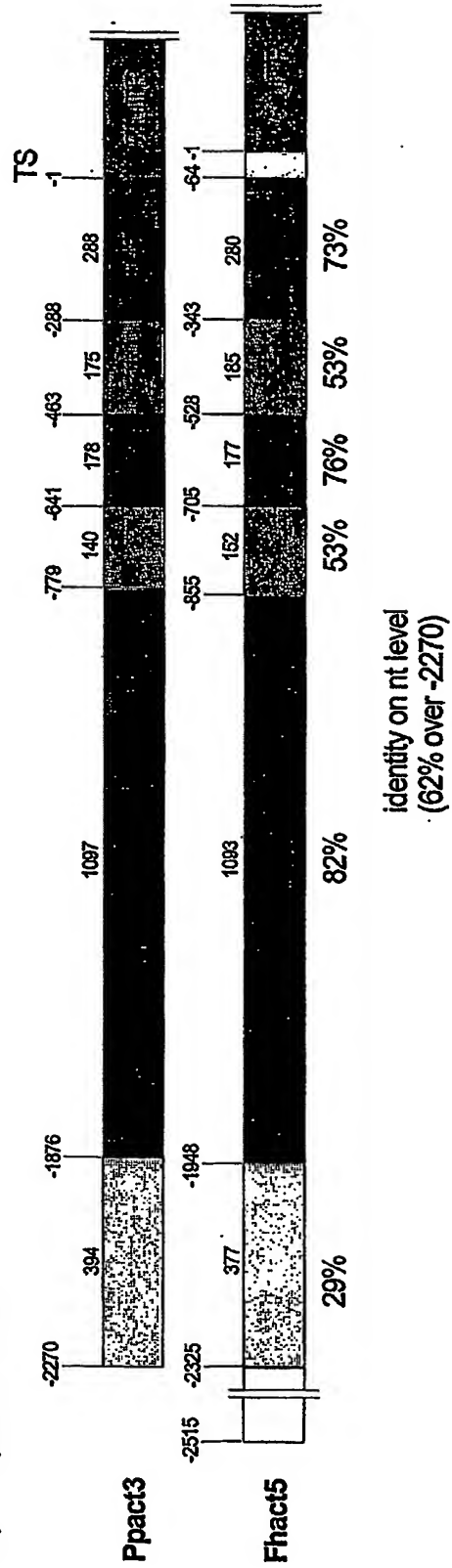
**Fig. 17: Ppact7 promoter:vegf deletion constructs.****1. 5' Ppact7 -1790:vegf****2. 5' Ppact7 -1070:vegf****3. 5' Ppact7 -854:vegf****4. 5' Ppact7 -659:vegf****5. 5' Ppact7 -484:vegf****6. 5' Ppact7 -299:vegf****7. 5' Ppact7 -66:vegf****8. 5' UTRact7:vegf**

**Fig.18****A) *Physcomitrella patens* actin genes****B) *Funaria hygrometrica* actin genes****C) *Marchantia polymorpha* actin genes**

A) Ppact1: Fhact1



B) Ppact3: Fhact5



**Fig.: 19** Comparison of promoter sequences of homologous actin genes from *Physcomitrella patens* and *Funaria hygrometrica*

# SEQUENCE LISTING

<110> greenovation Biotech GmbH  
 <120> Bryophyte expression promoting regions  
 <130> R 42095  
 <160> 27  
 <170> PatentIn version 3.1  
 <210> 1  
 <211> 1533  
 <212> DNA  
 <213> Physcomitrella patens

<400> 1  
 tagcataaga taaagatggt ctctacctaa tttattttta tttatcacta ataactcata 60  
 tcaatctaaa atatataaat gcctttaaca atagaagaat atgattcaac aaaccaatt 120  
 ctatcattaa aaatatatct aagattagat atgataaaaa tagataataa tattaataaa 180  
 tcattttaag gttgtaatgc aactataata atttttaata ttataacttt ttagtttttt 240  
 aaaataaaaa taaaatggtt aaatattata aaataattat actttatata tttatgatca 300  
 agttagtaca ttgatacatt taaagtccaa aataatttaa tgataccaac ttgcaaaaaa 360  
 tttaatatata ttaaaatatt ttaaaaagtt aagagcaaga aaaattattc taaatagaat 420  
 tcataccatg gtattataaa gatacaaaga atcaatgtgt atttattttat ttacatata 480  
 ttaacttgcaa tatatggttt atactacaaa tgactatata ttgaagatac taaccacaaa 540  
 aataaaaatc cagcactaga taattctaaa aacatgaaat acaataaaac attacattac 600  
 tagcttatat gggtactaaa tattttttaa ttatacaaat aaaaaataaa aataaaacaa 660  
 aaaaatccta tagtgacaag aaataaaaata aaataaaaaa attataattg accaatccct 720  
 aaacattaa tatttaaggg atattcatat gacaataaag ataatttatt tcatggaacc 780  
 ttgattatth tatcttttaa aggtggtatt tttaaaattg tttaatggta cttaaaatat 840  
 tgtatttata tagagaaaat cctccaaaaa aattctctca caagggaata gaattcctca 900  
 agtttttctc ttgactaaat tgaccaacca ccaaacaacc cacgtcatcc atccatccaa 960  
 cccccacaca acccaattgt ttctccattg tagacatcga caaatgaaaa tcatccgatg 1020  
 acgtatacac ttcatcctct ggtccctcca ggggtgccatg agccacatcc cgaccgccta 1080  
 tttcagatcc gacggcacag ggtgacagag cagcgggtctc agaccacgcc atttggaact 1140  
 cgccagccct gcccagcta acagtttcaa agctgcccgc cataaccggg tctctccagg 1200  
 gccgttagat cgtccatcct acgggagcac atataatact gccctagtgc cctaataccga 1260  
 tgggaacggg gagtccttta tctctctcgg aaagcgactc attcgccagt gtgcgcacgc 1320  
 cccgtgtccc aaggcaccgg gccagactct cgcacgggt ctaccacacac tcacccccac 1380

tcaccotgtg ttttctctgc ccccttcgcg ctcttcgtgt gtgtgtgttt tttcacggtc	1440
gattggcgag ttgcgaagga gggcaagggt gctgtgggtgc agcatcagct ggtagtaagt	1500
cagtcagggt tcgggtcgcg gtagttggac aag	1533

<210> 2  
 <211> 1539  
 <212> DNA  
 <213> *Physcomitrella patens*

<400> 2	
atgtatttcg gagcgatttc gtgtgctgtt ggtgtctttt ggttggaagc gatttaaaca	60
ggagagtctg tttggtggct tagggtaatt cgggtggagcc tgaaagatat tgctacgtct	120
tgaaatacca tcttgtttca gtgcgcattg cttgcaaaag cattgatagt tgtagcggga	180
tatggtgctg tttatggttg tatttgagca tatgtttcgt gacatctgtg ttgcttggtg	240
ggcttgccat actggttagt tcttgttgag tatcatattt actttccaat gtaatattca	300
acattttctc ctagcattac tataccattt ccatctatcc ccaatggcgc tategtctcc	360
ctgggataca ttttaaccat atttgtagtc cagtgcatta aatgcatgtg aaatcgcatt	420
tatagatgcg catattttaat gtcaaattag acatcttcac tcatataata cattttacca	480
aaaaatgaaa tgtacacaca gaatattttc aaactgccga ctatctcaaa aacctataca	540
ttatcaatct cattgacata cctcattgaa atactcctca ttgaaatact acataatttt	600
cattgtcaat attgccaaca ttcaaccatg agaagctgat tattatttct tttatactgc	660
ttactctctt aatgcaaatt caccattcct catgagagca gctgtatcta ctcccctgat	720
caatattact actaacttct caggaatagt actcgatatg ttgcgctggg tcagttacgc	780
aattataaag tccatcgtgt aaaccataat cgtcacaact ggatatctga tgccagaatt	840
tcagcaaatt ttagtgccga tccgaccagt tcaatgcaga agaggaatat aactatctag	900
aggttggttca caatcttttt cattacagtg cagccaaagt tctgcaacga agatacatc	960
gcaacttgca tgcaagggtga agacacatat cgcggctaga tcctcagttc gttgttaata	1020
cctgggaaag aaaatcaaca aatcgaattt ttctgcatca aatagccatg acaaagatta	1080
gtacttccag tgcaagtata gtctgcggaa atatatcgca gtcctcgtac tacagcttca	1140
aaatttgggt acatgacgag gatttcgacg cacaagaaca gaattaaccc gatcgtatcg	1200
agcgttacag taaacgagac gaagtgcctg tgtcttcaaa ccggcagatc tctacgaagt	1260
aagaatctac tcagcagtga gagcgagagc atctgggtgtg gcagaatcta ctatgattac	1320
aagtgccta tactgaatgt agaagcctgt atccacctca tataggaaac gaagtaatcc	1380
ataccgacat gttacatatc tccactgaag cagttccgta tgggcataca ggaaatgatg	1440
aagcacaacg cgtataccaa ttttttatca gatacacaat caccacaattc aaaacgcacg	1500

tttatacaac caacacgccca tactcaagat gagcaggca

1539

<210> 3  
<211> 1197  
<212> DNA  
<213> *Physcomitrella patens*

<400> 3  
tccttagtcg agaaggcgcg ggacgtgagt gagctctgaa gataagcttc caatttgcca 60  
ctgcaagtgt aacctgctcc atcgggcgcg agtccgtagg gatcatgaac acctcatttc 120  
acttggcggt agtgcaactct agcggcattg aagcaatcca tgccctcaga atgagtcgcg 180  
gggggcagtg aacgaactag ttaagaaatc cagtaatgac ggcaccacat cggcagatcc 240  
agatccattg cagattatcc tcttcagccg gaccgaataa accatgccta aataaccacc 300  
ggaatgtgtc ctgtgcgggg ctgattgttt tccaaagaaa cactaactaa ttatatccag 360  
acagtgggat gtatgcgggg atccgtgaag ccagatatga gatctctgat aaacctgagg 420  
aagatgtctt acatggcggc acgggaaaca cgaagaaaag ccgaggagaa ggtattgaaa 480  
gctgagcata gccattggct ggtgaggaaa gggcatgcaa caactcatcg aaagcggagt 540  
aaactttgaa atcccgtagg ctcatgcca tggtctaaat tcttagcctc gacgacgatt 600  
tcaaggtctg attcgaagct tccgagcggg gctccggaac tgtcacttca gtcgactttg 660  
aaatgtgaag cgactttgct cacttgtgac acagcaattc aactocacaa tataaaaaaa 720  
tcgcgaaaca aaaaaaaaaa aaaaaaaatc tactttactc gtcgatgttc cactcgaaga 780  
caaacagctt taaagcggtt acctgtggta gagatagatt tcggcgaagg aattcaaatc 840  
cagcaaccct cccactcgta ccgcagacct tgagtttgaa cggttctgtt gctgtttgcg 900  
gtgagttcaa aactcgactg acctctctga aaccaaagt ttaccttgag ctgcccgaga 960  
atctccgaac gttcgatata agatccaacg gtctcaagaa attctccctc gaggaacacc 1020  
tatgcccagg ggcagggggg tcttttatct ttctccctct gccgcaatcc atttcattgt 1080  
gcttgacgga ctgtcatccc tccccttggt gccagtggta tccggagggt cccgcgacac 1140  
cttctgggtc cggaactaag gtctgttggt cctttcgtga ggtagagcac actgaag 1197

<210> 4  
<211> 1012  
<212> DNA  
<213> *Physcomitrella patens*

<400> 4  
atgcgacccg aaggatgagt acacgogttt tggttttacg ttactgactt ttagctcctc 60  
cattcacact gcaggccctg gtttactgtt gaaagcacgg ttataccctc cgtaaaactga 120  
acattctgtt tcagcgcgtc gtgtcttagt tgtcctttgg ttcacttttt agtttggaag 180

caagtcgttg tatagatgat acttagcaca tatagttgct gtcgatttgt ttttaagttca	240
gcattccgct gcctgaattt cagtaaatac cttgtccaac ttcgatgcaa tataagttgg	300
cttcagtatc cagtcttgcc ttactccttc attgcaatct tgggtggcggg ctggtgcgcc	360
tcgtccactt tcacgatgta cctcgtcagc ttgtttgaac acttcctttc tctactgag	420
tatggcggttg gcctcttttt ccaagctctg ttgatgtagg tcctaccttg tcaaaacatc	480
accacagag atttgacgac aatcgtaatt ttaatccgat tgtatggggg tctgtgcata	540
gtcaatatat taacgcccc cctctcactt accaacgtct gttaccaact ggacaataat	600
gcattcacia ccaaagtgc atttttgtat gagttggaaa tatcgaaaca gttagtgcc	660
gtaattcacg caaatagttg tgtcatggaa actttttttt aactttctgt tgtccaatca	720
tcgtgctgaa acatttagaa atgtggcaga cagttgcatt tgatgtatca actgctgtgg	780
tagtaacact tgttgaaaact gtaagataga catgccaaact ttctgggtgct atgtgcta	840
tgtttatata ttctgaaga atggtacaat tcaaatgaaa gtgggtggga gaattgat	900
cattgatagt ggaatagggt attgcaatca gtgagtcctt ttttcagggt agctaatt	960
ccttactgat tatccattga ccaccagtgt ggcttgtgga atgogtgaag tt	1012

<210> 5  
 <211> 1386  
 <212> DNA  
 <213> *Physcomitrella patens*

<400> 5	
ccgtgggact tagttgtctt caattcatta ggaaatctgt ttgagcctct ttccattcca	60
atcttctcga caaaataggt tttttcagtg actcataact tattgtgctt tgcaaaattc	120
ccactaatcc gaaatgtatg gtgtgatcac cgagctttta aattgattgt gtttgggcag	180
tctacgaaaa atccagacgt ggagccttcg aggaacaggt tgttcgcgca ccgctacttc	240
tgaacttcac aacgcgcgt ctatgtcgt ctaactcaga ggctataaca caagttagcg	300
atgtccatcc ctctagtctt catatttgca acattaggag gaggcacacg ctggctcgaga	360
tgcccgtgga actcttccag attgctacca tcaatgcact cgtagacaga tccaaaagtc	420
attccacatt attcaacatt aagggatccc caactgacca accaagagca ggtgctatga	480
gtggaacttg ttattttcca aatgagcgtc gactacatat gccaggcag aaggatatgc	540
cgaggatatc gggggggcag gcatgtgttt tgtgtaaagt acccccagg taagaacttt	600
taagcggcgg cactggattc agaaacagtg gacagatata tccattgcca atgtattgat	660
tggctggcga agaactgttg caaaccacga ccagcogtag gggcgtaaaa tttgaatcca	720
ctgttttaaat ttcaaatttc aaacctcgac ggagtttct ttagcttttc agatgggcgc	780
agaacgggta ggaaactgtc ccgtcgccc aatttgaatt taaaaataa atcaaaacgc	840

tagagcttcg attagtatgg gcttttttca ctcttctgtc caattctttt tgttttttac	900
ctcatgcaag gcggtcggct aaagtgaactt acagggagga atattactga gagcaagagt	960
tttaccacgt tgtaggatct ggagaaatcc aacgatgcta ggcctacgca acgagtgtga	1020
ttcaacgccca gctataatct cattcgtgcc gtcgatcccg ccatccaacg gcgcagacgc	1080
tttgcgtagg aattgtacct tgccctacgat tggaatttga ctggcagctc ttgagctgga	1140
atttacttgt ctgcctgaga aagttgaagc gtaagatgct cgatccaacg atgggcagaa	1200
agtgttcgtg ggcaggaacc aaagccctag ggcgggctcc tcctttttatc tatctctctg	1260
gcatactctc tctcagtgtg cccccaggga cgtcttcttc tctccttttg ttcagcgtct	1320
cagtgtctga gggacggttt gccgtctttg tttcttcgtt ctogttgtag atcatcctta	1380
gcgaag	1386

<210> 6  
 <211> 997  
 <212> DNA  
 <213> *Physcomitrella patens*

<400> 6	
ttgtgacctc tcctctcggt atcattacgt agcacgctac gaacaggaca ttctgtttca	60
gcgtctaggg tctttcattc agcatttaga accaaatcat tgtatagatt tcaccagca	120
taccaagtag ctattgattt gttgtgagtt cagcatgctg ctgtctgac cgaagattat	180
ttgtaattga ctgttatatt tgagcatttc tgttcaatca tgtgggtgtg gtttgaattt	240
taattagcag gcactgagtt ccgtgacctg aaaagaattt tctgagaata gccaggtag	300
ttgcttcctc ttttctgtgc ggggatattt cttccgaaat atgggttatc cagcgtctca	360
tccgcttctg ctctgtgcta tgtgaacatg aatgcaattg atattcttcc aacatccata	420
taactaatgc atacttcata agaaagcaga ccgtcacgga taatgggaga aacattttcc	480
agtcactctc gtgtccacat ttctctcaca cgctaaccat gttagtaaac cgcaaggact	540
gttattaagc aatgaatatg tctgaaaatc gtatgtgac tgttgtcaaa gtgtcatagt	600
acccgtcatc gccgcattgt gcaactgctg cagatccgca gttaaataccc gctaacgaaa	660
ggaagagaaa gatgagagaa gatgagattg tcaccgggag agaatacagac gcagtcatca	720
gtgatactat tcgacggacc taacctcgtc cgtaaaatgc aagaatttaa cgaggcagta	780
aaatcagctt aaaacctccc cgcacgctta acgtaaccat ggctgtgcta aacatccacc	840
aagagaggaa acaccgcaca tgaacaactc ttctgaacta cacgtgaagc agagattgag	900
gcgaaaagaa agccacagat cgctgctcct caagtggtag atttattttc ccttgaaca	960
aaaatggagg tgtggaggcc aagcagcaat ttcgac	997

<210> 7



<211> 624  
 <212> DNA  
 <213> Physcomitrella patens

<400> 7  
 ctcgagtgca gtagacgaca aaatggaagg atgcgaccag ggatgaacgg gaagagtatc 60  
 attaatgoga gacccttgga gttgaaggcc acgagtggga cagcgatgcc gagaaaaatt 120  
 tgaaaatcgc tcatcccaga caaaatatct gtgggccagc caggggtttcc cagccagctg 180  
 ctctgccgtg ccagccgtag atctgctcat ccgacggcca ctgcgccccca tcctggactt 240  
 gtaccctccg gcatttgga agtggtcagcc totccctgac gaacatttca cctcggctgc 300  
 ccgggaggcc aggagcgtca gatgggagat ctgaocggcg ggcgaggag agacctgaac 360  
 cggcgggcag gggaaacgatg tcgttgcttg ttcttctggc tgaggcgtcc atccccctta 420  
 cctccgtgt gtgtttcaaa ggccgatata tgcgcttccc ttgcggaccg agctctgtcc 480  
 cgctcgttta cttctctccc accgagcttc cgaggttggg cattcccacc cttccttctc 540  
 ttctcctctc cttctctgct cttcttctct gttgtctgcg gattaggtct tgtggtcttt 600  
 cgagcttcgc acagcttgag caag 624

<210> 8  
 <211> 1146  
 <212> DNA  
 <213> Physcomitrella patens

<400> 8  
 gcgcgcggtt ggctggaaga agagtcgaga agcgatgtgc ggcagcggca gcagcaggag 60  
 gggcaggcag tcaggtgcag cacgtcgcgtg gggatgatgc gagggacttt gccggttggc 120  
 tggggtacag aagcgagggg taaatatagt aagattacgc gcggcggaag gacgcgatgg 180  
 ccaacgaggt ggagggggtt gggcggtttt acgtgtacag tatgagactg aactgacgt 240  
 tgatcctgcg cgaaccaccg gggctagcgg tagtagatag ttggagcag agttcgggag 300  
 cgttggtgcg gataagctcc ggcgtttgac ccaggggtgc aaccgtagtt gcatgggggt 360  
 ggtgggggga ttgaaattgg aaccggactt ggagttgaga agttcggggt gtttttggag 420  
 gcagttgaaa gacgttttta agaagtttga gctgttgga atacattgtt accctgagct 480  
 taagcagtgt gtagtggcga tgtgtttaat tgtctgattc ctgtatgttg gtgtgtgcga 540  
 ggcgtgtgag tgcgtggttg tgtgcttgac gtggcggtta tgggccgtgc tgcggaatg 600  
 atttactgga ttatttggtc cattgggttc gtggactgga gacggtggat gttttagtg 660  
 cttgtgtgaa caaggcgggc atgcagatga tgggctcgca ataaagacag ggtcatgtcg 720  
 ggtattgccc agatgaaagt ctcttttggg gatgccgata cggaaaatgg aagttggtac 780  
 agtcgcacgt tcaggcgtca tgggttgcc tggagtttg cattggaaga gagagttgag 840  
 ggtgtcctgg atgatgtcca cgaggtgggtg tttgaatcga tgttgtgcga agtagacctg 900

agcacogatg tgtgacacdg gaatgggtgag tttgtgtcaa tgaactgtga gcggttttgat	960
tgaggcagac attccaaggg gatgggtttt cgggttttgtc ttttaaggct ggcgcctgcc	1020
tagcctcctt tgtccttcag cgcattgtttg cttgtgacgt ttgcgttggg attgttagta	1080
ttgggtctgga tggaaatttt atcgtttcta tgggcagcaa ctaagtgcgt cttgtcattc	1140
ccatgg	1146

<210> 9  
 <211> 2973  
 <212> DNA  
 <213> *Physcomitrella patens*

<400> 9	
ggatccattc aacggaggat aagtatgtag ggtgatactt aggctcattc attcattcaa	60
ggcgtattta attaaactact aaagaaaaaa aggggggttaa ttgggggtgat tgggttatgg	120
aatgaataaa tgaataaatg ggtccccccc ctccccttcc tttcccttcc ctgcattaca	180
tatatatata tatatatggc atgcgggtgct gaggggtgtgc atgtgggggg ggggggtgtgt	240
tgagagtgtc aacgggtgcc gccacactct ccggacccct tcccattttc ctttcccttc	300
ctgccctggt cctgtgccct gctcccaccc actttccatg cccttgaaca cttcctgata	360
aaggccctcc atccctccct ttccttctc aaccatttta attctatggc ttaaâcatct	420
aaatcattac attcttatgt actaaaattt tatttataga ttgataattt tcttttaatg	480
aattaagttt gaattttatc tatgttttag ttccacaaga tttgttttat ttattacatg	540
aaacttcaaa agggatttga atatattaaa aatttccatt tataaatgaa tattcgagtg	600
agtttaatta aaattatttt tagcgtatat atatatatat atatagatat ggataaaata	660
caattgaatt aacctagggt taatttttat aacaatgttg aagtgcctt catgtagtgt	720
gagtgaagg atgtatttgg atatggatgt acttcaaaaa aaacatgata aataattgca	780
tagtattaaa gtttatgcaa taaagaagct agaaatgact aaaaattatc acaagcttat	840
taactcaca acaaatcaat gatatttcat atcaagtga actgttaaca aaagaaagaa	900
ttacgtgtat atttcatgat catattcttt tgataattaa tggtagggta acactatgaa	960
cataaaatta ttgctctcta caatttatca aaagtataat aaaacaaaaa taaaacagaa	1020
atcataattt atgagtctct acagggattc actgtcaaat attgtaagta aagtgtgtac	1080
tattaattga ggggattgtg gtatgccatt ggaatacgtg gatcaaaagc tgaaacacaa	1140
gaattttgaa actcaaaatt acattaaaat gtttgaaaaa taaacacaaa atacaatttc	1200
ttcagaaaaa aaaaaaaaa accatcgtca ataatgacag tcaacaaagt cagcatgcat	1260
gacgagctca ttgtatttcc tccaaaaaaa aaaaaaaaaa gaagaaaaag tgggccctca	1320
gttaaatacag agaatgccac atgggtgatag gagaagagcc gatcataggt gatacgtggt	1380

catgggatca	tcgtttccat	gcgcggaat	agatcgaacc	cctctcagtg	tctgacgggt	1440
caacacgggt	gatcgggtgg	accacacctg	accagcccaa	caaacgcag	ggaggaagag	1500
gtggcaagta	agtaagtcce	acgtggatcc	gagacaaaac	gttgtagcaa	taatatacga	1560
agtgagaaaa	aaccacagag	cgggtggcag	tcacgaagtc	gcagacacaa	accgggctgc	1620
ttgacacggc	gacccgttcc	ctgttctgcc	gcccgttccg	tcgccatctt	tgtctcattc	1680
gcacaagggt	ccttttccag	tgccttctgc	gcgggtccca	ccctctccat	ctgacccggc	1740
ccgggctaac	cggttccgga	gcagatgatg	atcgacctgt	ctcgagggt	ccttttgtgc	1800
accgcgtggc	ttcgtgattg	ggccattggt	gctgtttgct	gtttgttgct	ctgctttctg	1860
tgtccgggcg	gcattcctga	gaggcgattt	gcattgcgcag	gctcgttgta	gagcagcagc	1920
agcgctgagg	gtctcgtcta	ggcttagtgt	gcttctatcc	ttcgtctgtg	tcgcctctgc	1980
ttcatcgtcg	cgtctcttcc	tcagggttaga	gcactttcaa	gtgttggcca	ggactgagta	2040
taggaaggag	ggtttattta	tttattttatt	tattttattta	tttttctggt	atttttattg	2100
ctggctgatg	tccatcttcc	gacgcgatcg	tcgttttttt	ttttttgttt	gtttgtttca	2160
ttgtgttgga	ggagtgttaag	atttaatcgg	atgcatagggt	tgtgtgtttt	gcattgcgttt	2220
agagcgttta	catgtgcgat	gcacgagctc	tgggtgctgtt	tagaggccac	tgatttagta	2280
gtttcttggt	cgagggggat	tagatcttgt	accgcaagat	gttgctccgg	ggttggtggtg	2340
gcgatggcgt	tttataatta	acatatagtt	caatgggtgat	gatttaatta	gcagtgggtgc	2400
atgagttagg	tacggatcgg	gcgattgtgg	atccggactc	gtgttcaaca	ataggctgga	2460
ttctcttcta	ttgcgattgg	ccagttctta	catgcaatcg	ggtacacgat	cgctgaagta	2520
gaacaaatta	aactcatcga	ctgaattttt	gccgtctcct	gaactgtcga	aatagagctt	2580
gaaaatttga	ttgatagtga	ttgttttagtt	ctctgcgaaa	tcgttctaca	taatctttaa	2640
attctgaatt	aatctcaatg	tattttgaca	tcagctgatc	gcttgtccgc	tcgctcagtt	2700
caattcgatt	gagtattgcc	tgcagatttt	tcagaaaaat	ttaagtaatt	tgatagtaag	2760
aacttgactt	cctgtggatt	ttaaacagta	tagcatatga	agtgccagggt	tttctgaatc	2820
ctccatttct	tctaactcgt	atttccgaag	acttctatac	agtatggagg	gcgttctgta	2880
ctgtcctgat	tgcgagacat	gttttacgac	gaaaattttac	tgctccttag	aactaaaatc	2940
ttctgaaatg	gttgggcagg	tcggtattaa	gaa			2973

<210> 10  
 <211> 1128  
 <212> DNA  
 <213> *Physcomitrella patens*

<400> 10	
agcagtgcga	cacatctttt gcttttttca gcacgtctct tagctcggct tattgaactt
	60



gcagatcgtc	ggaaactttc	ttaagatttt	agatagaaag	gagacggctg	acaagaagga	660
gttgaaacac	aaattttgga	gagtgtttga	gtgatgagtg	atactgggat	ccttttttat	720
gggaaagatt	gccagcagca	gtaagcttgc	ttttgttaga	ttcctctccc	tacagcgtgt	780
acctcctcga	atatgcactc	aagcaagcct	agaggttgct	gctatagatt	tctcggttaag	840
acaggggtatt	attgaggcat	tttttgcgct	tccagatgga	gctactacca	caagtatcta	900
tcctattatt	atctttaact	tcgatggatt	tgccatgatc	actgaggtag	gtcgaagttg	960
tgattggact	tgtagtgatc	acttccagag	cgagctatca	aactgggtgc	tagaggagca	1020
acgcaaggag	tgctgaatta	ttctaataat	ctcatttagc	ctaagttttc	cgtcaaaccat	1080
agtgatgttt	ttaagttcat	ctcgtttagt	aaacatctca	aagaaggtag	accattaaat	1140
tattgcaggg	gttgtgatga	ctttatttaa	tagttgacct	cttcaattga	gaacgcgttg	1200
ctctcctttt	gtatagtttc	aatcatatca	aagctctatt	tgttctctgt	accttaagcc	1260
ttgtgtaagg	catttaaata	atctcttcca	cgattaagat	ggtagttatg	tcgccgggtg	1320
caacttccaa	gatgtcctaa	tgctatagtt	ctcattcaca	actcaggagg	tttgttgttt	1380
tatgtttttg	aaagtgcga	aggaaattgt	ttacttttcg	ctttgtgtct	gtgtatttta	1440
gaatagtacc	ttaacttctt	acacaatggg	gtctaatttg	ttattcttgt	gtatcacgag	1500
cgttaatcgg	tttgacgtc	ggaccctttt	aaaccaatctc	aattgcttct	gttctaatec	1560
acgcgtccca	cgaatggcag	gtcaaatacc	gattattgcc	cgactctaata	cgtgacagtc	1620
actgagacta	ataacgggag	gtcactatct	tgtgaogttc	tcgttatttt	aaaatctgta	1680
taatggcaat	ccctttctgc	accacggcga	actcatgatg	attcttatcg	agtcctgctc	1740
accaacttta	tcacaagacc	ctacggatct	aactatgatg	accaaagct	tgttctacgc	1800
atgcatgagt	cccttcgttt	gggagatttt	agaattctta	ggaactcaca	cgttgtccat	1860
aaattttaac	caccgggcaa	cataggatgt	tgacatgtag	tcacaaattt	agaaaaaccg	1920
acttcaaaaag	gttgcccacg	tagacaaaac	aactcgaacg	cagaaatcca	ggcgaccggt	1980
gaaattggaa	cattcacaaac	aaagcgagaa	gaggttcaaa	aaaaccgcag	agtaaaccct	2040
atgcgccaga	ggggaatggg	agatccacgg	gattcggaga	tgaaaaggca	tcgcgcgagt	2100
aaaaacaaag	agtgcgggga	gcaagggcat	ccagaagagt	ttcactgaga	tctacagtgt	2160
aactcagaaa	gggagccact	ggtacaaatg	ccagctttgc	aacgcagaaac	gaacgcggga	2220
gagctaacag	atccgggctc	aaaatctcct	tcttctacct	ctcaagccgt	ccacaaccct	2280
cattctccat	tctcgcaacta	ttctcctcaa	accagttgca	tctgcggttc	cctccatctc	2340
caaccctacg	gctttcgtgc	gagcttattt	gttgccctata	ctaagggttaa	accactcac	2400
tttgttgcct	atactttgct	ttgctatttg	gttgctttcg	tcttcgcttt	tgttctttgg	2460

tttatctcaa gtgcacatgt tctcgcgacg ctgtgccgct gtaggggctg gtgggcttat	2520
agacctgagc accgaggcgt gggtttgott cgactggctg tggttgtag caagggtgttc	2580
tcgtaaggta gttgtgttca gagctagatc ttgtgacggt gatgcgaaaa atgcgttcat	2640
cagagttaag tgatagaggg gcttttcgtg agatctgctt ctgtgatgga tctgctgtga	2700
aagcgggtccg cgttctcctt tatcttcagc tctgtgtctg atgtttggga aatgcacacct	2760
ttggatacgg tgcgattcag gctgtatatt gaatccccga gttttggaaa tctttatgac	2820
ctcacttaat ccgaaagcta atgggctgta ttgagtgagg ctaatacaca tctctccata	2880
ccgcgcttcg gtttcgactc gtcttacoga ccacattgat tcacatgcgg agacatcagt	2940
gttggatcac ttacagtctg acctaaatag cacgtgtgct acacatagtt tcaatgccag	3000
taacagtctt ttgatgtgca gagtatttct tctcc	3035

<210> 12  
 <211> 1221  
 <212> DNA  
 <213> *Physcomitrella patens*

<400> 12	
gctagtgcac acctgtctcc tgaaatgcta tcacaccttg tcagggtgggg ttatggagtt	60
tatttgtagt agctaagcag ctggaagagg ccagtgagag actgattttt caggggttgc	120
aagggaaatgg ttactcgagt aaagagccag cgctgtcgag accttcttgg tgcaattcca	180
tctttgaaag tatgcatcac aagttagatt cgtggctttt gagcttgtcc tcattatttt	240
gcctaccatt tatgtttttg tggatttagc atccgcggcg tttaagtttt tgttttaaca	300
ttctttcttg taggttcgga tagaatgttg gggacatttt atgcttgaag agcgtcttgc	360
actgtcggac tgtaatgcaa tgcttgtgga cctcagcctg gcctgcaata cttgtatatt	420
cgtgaaaaca atcatagcga ctctgtgtta ttcttcccat gtcattcact ggctctcgaa	480
ctttgtcgaa tacatctgat gggcacgcgt gcagaagccg ttctttaacc tcgatgggat	540
ggattagtac gatttgctgt catttaaac tatttgctat ccgtatttgt cttctgttcg	600
gaaatttggt tagcttgta ttttatggta tggtgtagga aatcagcttt ggtgagaaat	660
ttgtttcata acgacacaat ggaatgatga attaaattgt tgccagacca atatcgtatg	720
tgtcaatctg attcctcaat gcagatatgg ttgtggagcg tctgctgtac ctcttggtt	780
taaccgccgt atctgaacca actcgaacgt agtttgaaaa atgcactaaa tgatgcatat	840
tcaatcgggc aagtcataat aaacacgcgg ttttgaaagg tagcagggtg atataatata	900
aacatgtata tcgcaaaggc ccattcctga cattggatgg tgctaattaa gatctaata	960
accgttcctg gcaatgtatc tatcaagcaa actgaagaca caatgaatcg ttgagtgtat	1020
gtagaaacac aaaacgatct tgtatttctt tttcatgtgc cagagtgagc ctcatcgatg	1080

tacactgata ggactcaact ttgatatttt ttgaagattc ttatgcoctga ataaggtact	1140
tggaatcata gttctttgtc tcatggctta acttgattaa gatttgggga tttggaacct	1200
ttgtaaggag gcaatgaatt c	1221

<210> 13  
 <211> 3060  
 <212> DNA  
 <213> *Physcomitrella patens*

<220>  
 <221> misc\_feature  
 <222> (1)..(2301)  
 <223> a, t, g or c

<400> 13	
agactctact aattgacaag tatgtgacta caaaaggcca caagactctc tctgcactat	60
aactataagg ctcataatttt ttgtccatgt agcttgtata tatatatata tatatatata	120
tgtatattta aatcaaaaata tttttattca aaaacaaaat acaataaaaa accaaaaaat	180
attttaaaaa taaataaaaa attattaata cttttatgaa gctattattc aaattttattt	240
ttaatttcta atttaagatt tattattttt tcttaaattt attaaacttt ggaatttatt	300
tttaaaataa ataacaataa aataatttat agtggtttta ttgataagta aaattaagag	360
ctaaatttgg atcattatta caaagttata atacttaaat atttattgag atatatttaa	420
atttaattaa tatttttttat taagttatat atatatatat atacacatat tatgaaatta	480
tttaaaagaa gttagtagac ttttaaatat tttttaccat gttttaattc tagtacaatg	540
tatttaaatt atcttattaa gttatggaaa agaagttagt aggttattaa atgttttgtt	600
agattgggtg taaagggttt atgataatct tgtatgataa ggttggttag catagtttat	660
tttgcttaat taaaaaaaat tacatcttgt tacatttaaa tttaaaaaat acatactata	720
cacatatctg tatttagatt gcttttaca tttttatctt tttgtttttt gcatatttca	780
aagaaagccc agcatgtgta taataatttg tataaccctt agaaattaat aatatttaag	840
taaataatnc ttattttataa ataaattact gtttggtttt taatncaaga atttaaaaga	900
cccaattggt tattccaaag taatagtagc ncattaataa aaatccttca aaaatgaaac	960
taaacaaacc aatgcatctc aaatgaaaag gagaagaatg atcttacata gacanccaca	1020
aggagggaca tgacaactta attagactat gggtttagga acatcaacca ttccctacta	1080
ccaaaaaagc ttacatgatt ttaaataaca caatattcct tgtgactttt gtgcattatt	1140
gaggatatcc atctatctag attttggaca atgttttact gcccaaattt caataagaac	1200
cattcacata ttttgaaaca catttgatac actctacatt catgtctaga gtatagggac	1260
ttgggtttta gattaggggt tcagattagg gcttgcaggg ttacagttaa aagttaggat	1320

taaagattta gatggagtct tgggtcagag agaaaaaagg atttggggta aagtttttat	1380
gaaagagaat catcgcccaa acaagtagcg ggactgctga atgccttttg caatgaatga	1440
aaatttatca acgtccgtca atatgtacaa gaccatcaca taatggcccc cctgaccaca	1500
atttgaaaaa cacacacttc ctgcctggaa ccagtaatac aagtcatgtt aggggagaga	1560
gagagagggg gagagagctg tagctgcgta taataagggc ctgcgagatt cagtgtctacg	1620
tcgtatggat acaccgtatc acttctgggtg tacaggttac taaatactac tcgacacggg	1680
gcggggccgat ctgcggaacg cgccggggcc atgtcccagg gccctaggcc cgccataattt	1740
ctctcgtcca ccggggccta cgcaaacttt cccttctcac tttcccagct cacgctctct	1800
gttcaacgca caacaacgcg tagccgagac ggggttcggag cacaagtc aacagcccgg	1860
cccgaaccgt gcccgctctgg cgcctatctc tctccgcctc tgggcccgtt tcgctcctgt	1920
ccttggtgtg tctgtctggc ccttcaccgc gcttcattgc ttcttcgacc gagagcctct	1980
tagctcgtc ttgttcacca ctgcgcggc actccgaccc cttgcatact ctcttctgog	2040
gtgctgtctt ctcccatct cctgcatcg tgccctgttg tgtttttttt taaaggtcag	2100
tccctctatc acgtcagtg ttcgcatttc cgtgaagtgc tcagggtttt ttttgcgtgog	2160
aactgtcggg ggagatgtgc tttttgtcgt gtttgatgtg tgtgcgggtg agcgatgggtg	2220
ggtttcttgg aggaggagg agagtcttat tttagtcttg ttgcccggtg tgctcggggc	2280
gcgaatgtgg gtttatggta ncgcacaggt ctgcgtttgc gatatgtgtg tagaaccctg	2340
tgccgagcga tcatcataat agtagtttct cgtttcggag gggctgggct tgtcaagtgg	2400
aacgcagagt cgtagttttg agagttccag acgcgcacgc cgcagctgta gtgagatgta	2460
gcttctcggg gtgttttagtc aagggttcgc ttttcgac tcggatcatg tttacgtcog	2520
tcctttaagc tggatctctt gttctttaca gaacttggtc atcgccctga ctaagttgct	2580
ccagtgttgg tctgaagacg acaagcctct ttctttcttg aatagtaaga agaggaattt	2640
aatctgaagg cttgttttgt acagtagttg gtcgtttatt ctttgatgtt taacttagcg	2700
tttcgttgta cttctactaa tgtactctt agcttggtcc gaggtatta tttaatgagt	2760
catgccctga agtcgggaac agcgggttgc acctacaatc atatggatat gaggattcgg	2820
gtcgagtatt aacttgtagt cctttgttca ttgtttttga ttgcggggtt tagctgggtgc	2880
aactgcctga atagcaogca ctgctttccc tgcgttcgaa tcgtcatcaa cattactatt	2940
gtgtaatcca catggctaca gctgctgtaa ggttctgcgt caagggcgtt cttcaagaaa	3000
taacctatgt cttccttgaa attaaatatt ggtggttgtt gtgcaggtcc gtattaaata	3060

<210> 14  
 <211> 4124  
 <212> DNA  
 <213> *Physcomitrella patens*



<400> 14

attgtccatg tgcactacta aacatttttc agcacactcc cttccccggg attgagctct	60
tgctgtgtag aactctcggt gcaagtatca gtgattgcag actttgactg gtgagcacag	120
attcaacaga ggttttatttc gcagatgact atggtttgta aaaatagcag atatctgggc	180
tcaattctaa cggctgggtat atgtcagtac ctataaactt aactgtttgt agctctagat	240
cgggtgtgga aagtcggga ccaattcttg tcccttttcg tattaataa agggatattt	300
atttcatata tcgtcttttc cttttgtcat cacatctcta tcctgtgcat atcatgggtg	360
tattctcagt cgtaatggtc tttcaagtgg aatgatggct ttgatgatgt gcacctgggt	420
gtgtctctgg gcgtcatggg cttcacatga gctgcggta cagatcacgt ccagcctcac	480
acaattaact aggcattgctt tccatttcct tctgacgtaa atgacaggct ctgacaacaa	540
tgcttggcac ttctgacgt gggaccgctc gattgggtgc gaagtcgagc aaaattctaa	600
cctccacaac tgggtatctg aatattctag cctcttcctg agaacagtgc cggtcgatct	660
cgaattacct cgtaatagtc gtcaggcatg tatgtatgtt taaaaatact ccatgcggct	720
aaattatttt ttaaaattta tctttggatt tgaaatgaat ttctaccttt ttttacttta	780
agttacgagc tgcgattcca actaatgaag ttttacatac taatcagaag aatgtcgttt	840
tttgaaatta acaggttaag tgttttgaag aattaaagta tgatgattcg tcttttttat	900
atcaaagtag ttttgaatga ttctgtcttg cattttttta atcttggaat gaattgcgtg	960
tatgtgacgt gtatggaaag atacaaatct catgtagtcg agtacaagac aattacacct	1020
cttatgttta tggttcattt gtacatagtc tacgttagct taaggtcacg gtgtgtgagt	1080
atagtatata tcattacctt atttgaagtc cagtaaagt tagttatgtt accatcgacc	1140
agttatcacc gatgttgctg agaagcaatg tgaatcttag gaaacgagt atatttgaac	1200
tggatattaa ttcattcgta atctataaac agacatgctc tactagcgtt aaaacataag	1260
ctacagcaca aaatgatcta aaaaaatgtc atcaatcata agctgtgtat aatacatccc	1320
atgaatatca acagtatgag tttgggtgtt tgtgcacacg taaaaacgaa ccctcgaatc	1380
gaaatgtgta ttactgaatt cacatgcaaa tgaattgttt ggatcattta ctgattaggt	1440
ctgtactcta ttaatgaaac atataataga tttaagactg tccagtcagt tttgaattaa	1500
gccttgggat ttgtggtctc ttctcttcg gccactaaaa gtttaattca cattgatgtg	1560
aaagaaaaag tcacaactca gccttcgctg tgttagaaaa gctgcacgtg tgaggacttc	1620
tcaggcagcc tttccttttt cagttgagtg tcgaagtagg agcacacgtc gtcggttaacc	1680
ggctacagga ggtgtgcact gtccctttac cggatgtggg aagtcaccct atcctgagta	1740
tggctcacac ccaacgttgc tactccatcg cacagacagt tccacatgat agaactgtcc	1800
gcgagaagcg tcactctcgt gcggtctcac ggcttctgtt gcggccgatt cagtgcaggg	1860

agtcgttttc	gagcttgcca	agtggctctc	ttgtcattcc	cctgcttctt	ccggcggcca	1920
ttttgatgca	gaattgcgaa	ttctgcagaa	tatgttgaga	actcgtcttg	ggggttttcg	1980
gatgaggagc	taaaacccta	gagggacgga	caattctgtg	gagcttgctt	gtaatcctgc	2040
agtacaatag	aataatagag	cgacatgtcg	acgctttcga	ctcatgctcg	cgtgtcgtca	2100
ctgtcatcag	tgtcgacagc	gtcgaatgtg	gtggcaaattg	tggctgtgag	gccgtgtatg	2160
atagtatctc	ttctgcgggt	tgcgagaggg	ttgtgctcta	ggaaggggtt	gatgtcgcgt	2220
ggacctctcc	gaagacattc	ttgtatgaag	agtgtttcag	taatgccgag	agcttctctc	2280
ggccaactgc	ctgacctga	acagggtggac	ttgtacatta	atgcgttgtc	ccagacgccg	2340
gacgccctgc	agggattgct	ttcgcggacc	gaggggctct	ttttcacatt	ggcggatgtt	2400
gctgtggcga	ctgatcccag	ccaggtcacc	gacgctgtag	tgcagaaaaca	ggacggaggc	2460
tggcttggag	gtgtctcgaa	ttctcttgag	atagctctta	ccgtaagctc	tttttatttt	2520
tattttttata	tatttttgtt	tcttttttga	actgtgaatt	gtgtatatattg	ttttcctctg	2580
aaattttctt	tcagaatcta	ggtggtaaaa	cattctgata	cttatgctta	ttgcacgggtt	2640
tatctaattt	actaagattt	agtgtgaatg	tgatgatata	attttactaa	aatttaagat	2700
ttttctaaaa	tttaattgca	gctagtgtta	tctttcgagt	cgatgctaaa	acattcctgt	2760
tgacacgatg	atcatgaaag	ttagatgtgg	cttaataaca	aatgcaggaa	ttaatgaatt	2820
ttattttattt	atttattttt	gcagtttttg	aaggatacca	ttgctaagct	aggcatacct	2880
tattcgtatg	gtttcgcaat	tattctttta	actattttag	tgaaggcagc	tacttatcct	2940
cttacaaaaa	agcaggtttg	ttgttctact	gattttctta	ttttgtgctt	tctttctttc	3000
actttttgcg	tacaaatcat	ttttgtgata	tactaattta	ttgtgtaaaa	ctaaaagaat	3060
tactatattt	ttcagctaaa	tatctgtcga	tgctctgtat	ttactcataa	gttttatggg	3120
tttaagatag	taccagaca	ggactgagtt	ccattggtag	gtcagtactc	ctgttagatt	3180
agggaggctt	ctattgttgt	atatctaatt	gaaagtgggt	atgtttaaca	ggtagaatca	3240
acttttagcta	tgcaaaaact	acaacctaa	ataaaagcta	ttcagactcg	ctatcagggg	3300
gatcaagagc	gcattcaatt	agagactgca	agattgtata	agcaggctgg	agtcaaccct	3360
ctcgcaggtg	caattttgtc	gaagtcctcg	aagcattaat	gttaagaatg	cttgacagtc	3420
actttccgggt	ttttgacgga	cacaaaatac	agtcgaaggg	actaatactc	aataacttgg	3480
ttctgtatgg	tagctcataa	gggttgtgggt	ttatgatttt	acagggtgtc	tgctactctt	3540
cgcaacccta	ccagtatgga	ttggattgta	tcgtgctcta	tcaaagtgtg	ctaatgaggt	3600
attgcatcat	gaactggagt	gcttgaaaaca	gttgccttg	tgcggcatgt	tgttccacct	3660
tagtttattg	tgaacatag	gcgtcattag	acaatccaca	tttagagtaa	tacaggaagg	3720

tcttaccata	táttcatttc	aaagaggttc	aacagacatc	gtaatgcaaa	gttctgtaca	3780
ttttctcttg	acttcaacgg	gagaatatct	attcttaaat	gagatatttt	ctgtgggtact	3840
ggtattcaag	tatgaatgta	tgtaactatg	atttacttat	gcagttctgg	ctttgcaggg	3900
gctcttgact	gaggggttct	tctggattcc	atccttggca	ggccctacaa	cgattgctgc	3960
tcgttccagt	gggagtggca	tttcgtggct	atttcccttt	gtggtttagtt	agtcccttca	4020
gatgcttgtc	ttcgttáttt	tttttccata	tcaaatgtaa	tgatgctggt	catacagtaa	4080
catatagtga	atttgttgat	caaaatgggt	gtccatggaa	gctt		4124

<210> 15  
 <211> 3053  
 <212> DNA  
 <213> *Physcomitrella patens*

<400> 15	
ttgttgaatc	atgttaattg ccaatggtta ttaatgacca tcatattgta cctggaatgc 60
attgaaaag	taatgttcca ctaaaataaaa gttgatccac caaatattgt tgtctagtca 120
tatcgacaaa	tagattcaaa ataaattaaa attaaaattg aaaatgtata aacattggca 180
tgaaaatgat	attaatttaa aacaattcaa aacttataca attattttaa atacattagt 240
caccgggtta	aaggagacag actgacagaa ttggattgcg gcaatcagta gcactgcaca 300
aataaattta	acatgaaaac attatgattg ctaatactct gtttgcacgc acttctacaa 360
caacaaaaac	aaaaaatata atcaaacaaa acaagcaaac aataaatgat tttagatttt 420
gcatgatata	agcaccagag ataattatga ccatgtgata aatacaattt ggaccattta 480
tatcctacaa	aaaaaagaaa aaagaaaaaa gaaaagtttt tgtttgtatt tgatatcttt 540
attttgttac	caaaattaga taattgcaag ccttgatttg tctgagatgg aatgtatatg 600
taacacattt	gagcaaaaaa tttaaattaaa tttaaattaaa taagattttt ttatatatag 660
taaattgtaa	aattgaccca aacatttact .aatcaaccc acccattcta accatcataa 720
gaagaattcc	gctatcaaatt ccaggttggg tgaaaaccaa tgaaaaaatg gttggcttct 780
caaccaatga	taatggatgg gttaatttaa taaattcatg ggtcaattta aaaattccat 840
atatataatt	aaaaatcaat tgcaaaaaat attttgacac aatcacacgt gttttgaaaa 900
tcatacatgg	acaaaaatac aaagagattt tttaaccaat attttggaac cacatttagc 960
aagggtgtcca	atgcccttcg ataccacaaa gaacacacct tactttgccc atatttaccg 1020
atatatgctg	cagtcagtta ggggtgaatc cctgagggag gggggctccc gtgtgaacaa 1080
agtccaatgt	ggggccgccc aggattaggg caccaggtgt gaacgaggct ccacccgagc 1140
gagagccagg	aatttgaaac tggcatggga aagggggttg gttccacctg atggcacctg 1200
cccaccacca	ctagtaaaga ttcaatgccc accacactgg tttttgaata taggatcttc 1260

cttctccttc taattcttct cttgatggat gaataatata accgatgaat gagtgggcac	1320
atggacgggc ctgcgccct ctctactctc tgcaatacat tacaaaatac atacatgtat	1380
acatagggat ttgatgactt caatacat acactacaaa accgggtcag gaggggggta	1440
taaccaggca agcccgagtg gcgggcagta acaaatacac acccccaaat cgtatgggcc	1500
ggacacgtct gagcgacacg cgggtgccct gccctcctgc cccttcctc gcccttttc	1560
tctcgaccgc ctgtcgccgg ccgggccag actcctgcca acctgggaac caacccccct	1620
ttttggtgag tgctcttcac ttccctcgca ctcgctgctc aagttgaggg agggagggag	1680
taggagtagt cactcaccgc gcctggcccg gtccggttcc ggtccgcggg ggctgcgttg	1740
cgcgaccgt tctcgtgggg ttatctctgg ttctctatcg ctcgctcttg tgcctcgtac	1800
tgctcctact ttttccatt gttgctatgc tcgctgccct gcgctgcttg gccgtccgtt	1860
gtgcccctcg ctcgtcaacc aagcactgca gtctcctccc gcattccttt ctgcagcagc	1920
gtgtatctct ctctctctct ctctctctcc tcctctgttt agcgtgggtg ccggttctct	1980
taagggtgaga gcttctgttc tatcgggtgt ctcggttttg gtatgtgtgg tgaccgacga	2040
tcggtttgtt gtgcacggtc gctggatgta tggctgctct tgttcttggt tagttctgtg	2100
tggcgattaa cgtgttcttg gaggagtatt tttggccttt gtctgctgat gcgctcagca	2160
gcgttgcggt agtgtaggct tgtgcttcac atgagcgtgc cgcgcgtcta ggcgtgggtg	2220
ttgagttgaa tcttttgccg aatgactata gttattgatt tcttgttatc tgaagatctt	2280
gtgctgagat atgtgggtga gggattcgag aagtgcctac cccttgttgt gatgaacagt	2340
tttcatttga tgtggttacc atactttgga gccttgcatc ccggatcgtc attagcttca	2400
tctacgtggc tggatttttc cgtcaaccgt aggctgaagt gccttaaggg gttacatgtg	2460
ctgagttgac tacatgtaac aatggcatgc aaactgattg cgtgcacttc atacttgtat	2520
tcagttcgtt gtagagtcgc ggatataat taggtagaat aaagaatctt atctctcggc	2580
attcgaataa aaatttcac ctttttgaat gcaccttggt tgaaagggtc ccccatgccc	2640
acggttgact gagaacaatg tctgcgcac agttactgat ggtcgacact gttgtcacta	2700
atttgagtga ttaaggtttc ctaccggctt tttcttttcc actgatttag tttattcttc	2760
atcaagttta caaatattgc tctgtatata acggtttttg ttagtctttg atgtaatcat	2820
attacctggg tttattatct agtgaactat gactgatatg ctggcgcata ttctcctact	2880
taatttgacc ttattagaag atgttcgtac ttagagtacc ttttacttaa tgtaactgaa	2940
tctatcattg ctttctgtct taatcgtgct acaaaattta actcattctc tcgttaacta	3000
atgtttttga gcacttgcac tgtttttgaa ctctgtagg atcattctaa aaa	3053

<210> 16  
<211> 1879

<212> DNA

<213> *Physcomitrella patens*

<400> 16

atctgtactg cacagtttta catttttcag gcttgcattt tgctgggatt gagttcttgt	60
tttgatagaa ctctggacgc aaatgtcttt gactgcttag ttgggctggc gagcacacag	120
taagaagtgg tacatgttgc cgaaactatg gatttgtaaa aatgaaacgt atctgggcgc	180
ataacgaact gcttatatat gtcgctgtct gttaacttca atctctacat gtccagatcg	240
atgcggtaga acccgaccat tttttgatcg atgtttgaac ctttttatgt taaataaaaag	300
gtaccatggt ttcagcgc ataatcatatt tattttggtc actatggact tgatgtacac	360
cggatgttac agctcagttc tacttcacag ttattcactg acttgccctg aaaaagtcgg	420
agtgcagatc tcgttgtgtt ttggtaattc ggttggccag tctcagagct ctattttttg	480
atgaatccag ttgattggca ctcaatgttt ttttttattt tttactttta tcatagtgtc	540
aagggttgcta cgccaggaat gctgtgaggc acattctacc cgtatgaatt tcctcgttcg	600
caatagctgc aagctcaatt taggtttttc tgagcaagtt gtagaactat cgtgtactct	660
caccagatth cagcctctca gtgctgagt ctttcgtcac gttaactaat tgtggaagat	720
ttggaatcat ggttgcattc cttagtttga cagaattcac agtcgttagt tgacctctct	780
atcttgggtc accatatgtc aacctgttca agagggtgt gctcgggttag gtaatcactc	840
agaagtttct tcctacagaa aacttgtttt gtgggcatca tctacgtgga agaattgttt	900
gagcattaaa tcattcaaca ccttcagtta catgaagtag gttggaagca gtgccttgaa	960
gagatccttc acagaaagcc totcaattct catgaagtct gcatctaact tcttttgaag	1020
tttgtacacg tgtgggcaga attgaagttg gttttgtgtt gtttgaaaca actgtaattt	1080
aataaatccc aaacaagact aaggccatct aacgttttca catgttttta aaaattacat	1140
tgaacttttg gctaccgtag ttttagacag atgcaattaa aaataaaaag aaaaaaatga	1200
aaagaaaaaa gtcttgtttg ttttagttgt ctgttttgta cagttttgtg acctatttta	1260
gagtgtcatg tatcgaacat ttgactcaca attataaggt tttatatattt aaatgagtct	1320
tgttgtcttt tattttattt tgttctacat tctgtaatat taaaacttct attgaaaaca	1380
caacaaacat ttaatttcaa gtttttcaaa tttatatatg catattttgt atgtaaattg	1440
tacaaatggt cataatgcaa attgaaatat ttaatgtaag attatagcac ttaaacctga	1500
tccaaaagat aataattttg ggcaaataat taaaattatg atagacaaag tttagaatgt	1560
tgtaataaaa atttatggta agtgctaaag tatgtaaaac aaatttcata aagaattgct	1620
tgtagcattt tcaagagaaa aaaataaata cttacgacta tttttaaaat gacacaaata	1680
gtaaataaca atatattgat gaggatatat atatataatc aaaattaacc attagtgatt	1740
tttaacctgc atagtattaa tgtatgggac cgcaaggtag acacctacct ctactggata	1800

gcacctctca tatacacaat aaaactttta ccttgctaaa agtccaaggg aatttacaaa 1860  
 agaaattctt ttaaaaact 1879

<210> 17  
 <211> 1823  
 <212> DNA  
 <213> *Funaria hygrometrica*

<400> 17  
 ctttcgtggt gcctcaagag tgcctcgcga agaaagaagg ttccagcaac aactagagaa 60  
 tgggtacagc attcataaaa ctacagataa ttatccttca aataagtaag aaaaaagaag 120  
 gaaggaattg ataaataagc aagaaattaa gcaaagcagc cactcggcta gacaaaagag 180  
 actgcacacg ggtggccaag gaaagcgccg gtcatagggg atatgcggtc atgggggtcac 240  
 tgtttccggc agccggaatc gattgcaccc tcgcagtggc tgacgagtca gaaccgggtg 300  
 ccaagtggac ccagctcagt cgcgggcagg ccgaggtggc accgaagcct ggtcaacgtg 360  
 gaatggatac gaatgtactg gatacgagat acgaatacga tacagtagag aaagaacgcg 420  
 gcgaggggtg cacgaattcg cagacacaac cgagtcggcc tgacaaggcg ccccgctgt 480  
 totgcgcgcc cttccatcac ccgctttgtc tcattcatcc acggctcctt tttagtgtct 540  
 ctgcgcgggt cccacccct ctcactggac tcgagatgcc gccctgcgct gcctgactcc 600  
 acctggcccc gcccgaccg ccccgaccg ttccatggca gatgttgatc gccccgtctc 660  
 gcagctcctt ttgtgcaccg cgtggcttcg tacttggccca ttgttgctgt tgctgttgcc 720  
 ggtgctctgc totgtcttcg cgaggcactc ttgaggcgat tttttttgta gtagcgcaag 780  
 ctogttgtgg agccgcgccc agtaaatacat ctaggcttag tctgtatcca ctaccctccg 840  
 ctgcgatcac ccctgcttcg ttgtcggcgt ctattttctca ggttcgagtg tttctgagtg 900  
 ttggcgagga ttgagtgtag gagcgggagg ggtttgctgt tgtttttgtc gctggcggat 960  
 gtogatcttt cgacgcgatc gcatttttct tttgattgtt ctgttttgga gaacggaatc 1020  
 ttttgattgg atatatagat tgtgtgtttt gcatgcgttt agaacgttta cacgggcat 1080  
 gcatgagtc tggtgtcgtt tggaggccac ggatttagta gtttcttggt caaggtggct 1140  
 tagatcttgt actacgagat gtttctccat gattgtgggt gcgatgactt tgtatacttg 1200  
 acgtgtagtt taatggtgat gattcaatta tcagtgggtc atgattttgt tacggatcgg 1260  
 atgatcctgg atccctgatg attctttttc aagtaggttt aattctctgc aagcgcaac 1320  
 ggttggtcgt ctattctaa tgggtggcatg atcgcttatt aaattacgtc gactgaattt 1380  
 tctccgtctc ctgaattgtt ggagtagcgc ctggaaattt gttagatgga gatttttcca 1440  
 ttatccggga aattattcta ttaattcttt tagactcact cgctcataac gcatattgaa 1500  
 ataaaccaca gatgattgct tgatcactta ttcatttgaa tttgacagaa tacttccct 1560

tcctgtttcg	gtgaattaaa	ttatttcgat	atttagaatt	taatttaata	ttatttttac	1620
acagtacaac	gaatgcaaag	tggaggagtt	gtcaggacaa	ctgaatccct	cagtttttct	1680
agtctatatt	tctgaagact	tccacacaat	atagtagacg	ttctgtgcta	tcctgactgc	1740
aagacaaaat	ttacgacgca	aagtaacatc	tcctttttta	atctgagatc	tcttcaaatg	1800
gttgggcagg	tccgtattaa	gaa				1823

<210> 18  
 <211> 419  
 <212> DNA  
 <213> *Funaria hygrometrica*

<400> 18	
aggagtgtta	cacatctttt acttttttca gcacgcctct tcgctcggtt tattgaactt 60
cgattacaaa	cttgtgtggg taccgaacta ggccggctag cgtagatcga gtagaggctc 120
ttgttgacag	aagttttcgt ttgtaaaaaat agctgatatc tggacacata cgagtggctg 180
attggattca	gtgacattca cattatttgt taacaggctc agggttgttc gtagagtctg 240
gccccatttc	tcgtcggaat gttggcgccg ttttgtgtga aatgatggtg attatggtta 300
aaatgcatgc	gtagtcctgt tgactatggc tgaatggata agatatattt ccatcatagg 360
ttagatttca	agcggagcgt gaactgtgac gctcaatcac agaatgcgtc gtcttagcc 419

<210> 19  
 <211> 1333  
 <212> DNA  
 <213> *Funaria hygrometrica*

<400> 19	
ggatccgaga	ggaaagagag agaagaggga gcgactcatc tagccaggcc cggtcgggtc 60
ctctgccctg	cctggcgoga cccgttctcg tgcctatctg tggttctcta tcgctcttgt 120
gcctcgccct	gcacctcctt ttcccattgt tgctgcttcc tgccctgtgc tgcctggccg 180
ttcgttgtgc	ccctcacctg tacactctcg cagccaagca ctgcagtggc agttcgcctc 240
cgcattcctt	tcgtggcgcg gtatcccccc cgtcatcttt ttcgtcggtg acagttcttt 300
gaaggtaga	gcctctgtcc tgctgccgtt ctgctgtgc ttgtgttgtg gccgacgac 360
gggtttgttg	tgcaaggctg ctgtgcgcat cgtcttggtt agtattgtat gtogattact 420
gtgttgtagg	agcagtggct aagctttgtc cgtgatgtg gcaccaacg gcgtcgctca 480
agtgtaggct	ttttctttac acgagcttgg tccgcgttta tgggtgtttg atgttacttt 540
tttcccgaat	gacgatattg tgtgatttct ttacaacaag agattttgtg acgtgaactg 600
tagtttgtag	attcgaaaag tgttgtttcc tcgtttttga tggacattac ttatgccttt 660
tagttgtcac	ggttggtggc tttgcattct tggtcgtcat tagtttcac cgtatgctga 720

cattcgctac catcccaagc tgaagtgctg aagttgattt catatgttca gtttgctgtg	780
tgcaaccagta tgagtcaaaa ctgattggat gtccttcaca acttcattct cttcatctta	840
aagtcgagta caaatcaata ggtacaggac tcctatat ttggtgttccg ccatagttat	900
cgtctttcgt caaaattacc ttattgagag gacttttcct tgcaaaggtc tcatcgagac	960
caatctctca gagtcagata cctatggctg cagcagaaat ctctagtcaa tgtttctaag	1020
ctctcctaag gattttcgtc ctttcatcag atgtattcta tccaactcca agttcgcaac	1080
aattttctca tacatcattg tcttctggtc tttctgttct gatactgcac cgattcattt	1140
taggatctta taatccgtgc ttgatgtgcg gatatgtgaa ttccctgagt gttcacctca	1200
acgtactcaa agttgtttota ctttcagcat ctttcagcca atgcggcaga tgcgatcact	1260
tccgaggact ttaaaattct gtactgtttc tttaaaacgc ctttttcgat tctatgcagg	1320
atcattgtaa gcg	1333

<210> 20

<211> 3289

<212> DNA

<213> *Funaria hygrometrica*

<400> 20

atgcatggca aaacatcccc tgtcttccat gatgagaaag gcgaacctgg actgcttgat	60
ggctcttcca ggtatctcat tgtgcttcgg tagttgttga cgtcttcact tctgcttctt	120
tgccttcttc ttcttcttct tcttcttctt ctttctctct ctctctctct ctctccaaa	180
ccttcttctt gtcttcttc ctcttatttt cctatgtcaa tgaagtttag cacctcctaa	240
aatttttgga tgctgttttt taaatagaag ggacgggatc aaaggacgag tgagtgtcgg	300
cttttgcaatt gcttccgttt tataacaacc tattaaggac gtagatcgtg tctgtaaagt	360
catctcttat agccttttat agtcttttta agagagaaga gccacctctg agtttcttat	420
agattcggac aagagatgtg acgacttagg aagtgtcttt cggaattttt cttgtgataa	480
tggcgttgca tttcttgtcc tgtcttattt ttaactgaac agtatgtacc atttttccgt	540
atagtcctta ctttataata tgtcctcttt tctttcgcct cagttctatc atattctttg	600
atatgtacta ttaactttcg ctatctgttt tctttagtgc ctttcaccgc gtgccgctat	660
cacagcttgg tcatagagga ggcctcattt ccagctgacc aactcgagat tacagcatgg	720
actgaggacg ggcttgtgat ggggggtcgt caciaagtct acaagcacat ccaaggagtg	780
caatttcctc ctgagagcat cgggactcaa aatgggatgc agatcgttgg aaattttctc	840
aagatttttag atagaaaaga ggcggctgac aaggaaggag ctgaaatgaa aattttggag	900
agtgtttgag tgatgagttg tactgggtata tcttttcttg tgcaagattg ccagcatttg	960
tcagcttgct tttgttagag tctgacccc cagcgtataa ctcttgagt atatgccc	1020



gcaggcctag atgctgctgc aataaccttc tcggtgagac agggtagttt ttgagggtatt	1080
tttgcacttc cagatggagc tactactaca aatatctatc cttatcttác gttaaactac	1140
gatggaattg ccatgatcac tcaggtaagt ttaagttgtg attggacttt tagtgattac	1200
tttcagagcg agctatcaaa ctggtgcttg gaggagcaac gcaaggaatg ctgaattttt	1260
ctaatagatct aattcagctt aagtttttcg tcaaacttag tgatattttg aagttcatct	1320
cgttagtgaa acatctcaaa gaagtacgcc attaaattat tgcagggcct gtgatgacat	1380
tatttgatag ttacctctt aaactgagaa cgcattgctc tcctttgtat agttccagtc	1440
atttgaaagc tctatttgot ctctgtaact taagccttgt tcaaggcatt taaattccct	1500
cttccacgat aaaaatggta gttatgttg tggttggaac tttcaagata ccataacatt	1560
gtggttctca ttcacaacgc aggaagtttg ttgacctata tttttgaaag tggcgagtga	1620
aattgtttac tcatcacttt atgtgtgttt ctagtatgtc acttcaattc cttcctcaac	1680
tgtgcctaata ttttcatctc tgtgtgtcac gagcgtaatt tggccttagac gttggaacat	1740
tctaaggctc cagtaaccag ttttcattta ttatttttaa attcacagcg cctcaagtaa	1800
tgaaggaca aacgcgagtc attgcgcaac tctaattgtg acggtcttca agacaactaa	1860
cggcaggtca ctctcttggtg atgttctcgt tgttgtcaaa cctgtataat ggcaattcat	1920
ttcgacatca cggcaaactc atgatggttt ttaacgtgat ttgctcacca ctttcttcc	1980
aaagttatca ccgacaccct atgggtttta ccatgttatc tgaaagcttt ctctacgtat	2040
gtatgaatct gctcattagg gtgaatttgg aacttaaaga atctcacacg atgtccatga	2100
attttgttac tggacaacat atactgttga ccacatagat atgcatgttt agaactgcaa	2160
aaaagtttgt tcacgaagac agaacgacta gaacgcagaa tacctgcgat cgggtggaatg	2220
ggatcatttg cagtaaagct agtaaaggat cgaaatagac gcagagtaaa cccgatgcgt	2280
tagaggggaa tgggagatcc acaggactcg gagagaaaat gcaaccctgc gggtaaaaat	2340
agagaacgcg aggaggaagg gtagccagaa gagtttcacc gggatctaca gtataagccg	2400
caaagggagc cacgggtact agtgccagct ttgcagcaga gagcgaacgc gagggagcga	2460
acagatccgg gccccaaatc cccttcttct atctctcaag ccgtccacag ctttcttct	2520
ccatcctcgc actattctcc tcacagcagt tgcatttgtg gttctctcca tcttcaaccc	2580
ttcgactttg gtgcaagccc gcttgttatc tatcccaagg tttcacgcac tcccccttc	2640
gctgtgtgtt tcgttgcaat atttttggct ttagttttta ggtttataca tagtgcacat	2700
gctctcgcaa aaccgtgccg cttcagggga tcgtggttct gtagacttga gcacagagat	2760
gcgggtgaac tottagtggt cgccgctgca tccccagagt agttatgcta cctaaagaag	2820
cgtgctcgta cggtcgatat gtttagagat ggatatttag acgatggtgc gtgtcctgcg	2880
gtcatcagag taggtgaagg gatttttcgt aagatctgct tttgtgacgg atctgcaatg	2940

caggaggtct gcgtctttct tttcttcag cttcgtgccc aatgcgtcaa atgcgcaccc	3000
attgcacaga gtgctattaa ggcggcttca tgaagctccc agttttgtga atcatgttaa	3060
cttgtccact gatcagaacg ttcgggctgg catacgtgaa gcgaatacac atttttctac	3120
agcatgttcc ttatttttagt cttcatactc actgcttcga ttgccggagg gcctccatgt	3180
tcgaccacat cttcacacgg ggcttatcat ctgacctaaa tcgcacgtgg cctctgtatt	3240
gtgtcaatgc cagtaacagt ctttttgatg cgcagaacat ttcatotcc	3289

<210> 21  
 <211> 937  
 <212> DNA  
 <213> Marchantia polymorpha

<400> 21	
catatgcgta cggagttgtg gtccccgac gccgtagttg ctgttggtgt ctggtcacag	60
aggattcttt gcttcgcttc ctaatgtagg tggccagggg tggatcgtct tcctcctacg	120
cttcgtttgg acacatacat ctggatcttg agaggaacac gtgaattaga gttacatgcg	180
gtattgcgtc atctttgcga ggtaacggcc gcgccgcaga cctagcggtt gcttctgcgc	240
gactcaagga atcttccctc tcctgctcca tcaactggaat gagagttgca gtctgatott	300
tgggaaatct ttcattcttg tgaccatcga ctctgtcctc tcgatgaggt ctgggatgat	360
tctgcatgtg ataactagcg agtcttcatg attgtcacat gcacccagat gcgacatctg	420
gcgcgctttg tgcttggtca tagccgcctt cttttatott gatttgecta atgagcccca	480
tttccagacg tggacggcag atcggtcata aggtccaaga gcaggaaatg ctatgaggcc	540
gtttgcgtgg tctacctctg ctggcctgcg aaaagactgc ctgtccgact tcaatatott	600
taaacattag gctcttcagt tgtctcgctc agaccattat tatgagttat tgttaccgta	660
gtgtgttgct atgtcagccc gtgtagtctc gtcaatttct ggagggtaat gcgaacttgt	720
tcatgacggc acgtatctcg tcgccccgaa gatcaccctt gttgagaagg atttcatgcg	780
tctgcgtcct cgttcatgtt gacatgaatg atagaagccg ttctgaagac acgaaatgtg	840
gttgacatat acattgtgat gctcatgtct tttgtcgagt caccaagatc cgcaaccatc	900
tcatcttctt tcattttggt taggtaactt cgcgaaa	937

<210> 22  
 <211> 3025  
 <212> DNA  
 <213> Marchantia polymorpha

<400> 22	
tcatgatgtt aagcgttttc ataatacaaa gaggttttgt atatagataa aatttacttt	60
ctgaatatgc aagcatcata ttctaaattt aatcgaacat aattttttct gagctttctc	120

tttctttttc tttaaattaa atttccttca ctgcaatfff tttattacga ctcccacgag	180
gagtattttc cgactataga tcttagggta tataactata tatcacgctc gttctaaaca	240
ttttttctaa ttttatgaaa agagataaat atattaataa tatagggtat ttagattatt	300
gaaattcaca gaaaatacca tttttgtctc attcgatatg ttctagatgt gtgtgcgtat	360
atggtcatat acttgggata tttttaaatt gtgaatacaa gattataaca aagttatcat	420
tgcaaaatac taaagataag ttatctttgg tgagaagaca tgatatacca tctgcatatt	480
acttattcac caattgacca aagatttaca atctaccttg atgaaccata aatttgagaa	540
ttttatatgc agatatttgc ggatctttcc aatcattatc tagctcttgt ttacattttt	600
gctttcacia aaatgcaata atgtgaaagt tgatgcaata atcccttttag gttttttgac	660
tcataacaat tttctctcca aagcattgag attcaatgtg gacgtgatac ataaattcac	720
atcttgatta gttacatata aatgtggaac tgccgtatft gtcggaaagt tcatacaatt	780
ttttttgttc atttgaagat cataagatag ctgcatatat caccattagt gatgatatga	840
tatatgacat gagaaaaata taacttaata tgaaggaaagt cttgatatgc cttgctatcc	900
ctaggttggg gtaggtcttt ctttcatttg cgattattat tactgtgagg aatattcggt	960
agaatggatt ccttggaagt gttgtatfff tgacctctca taattaagca cagattaatc	1020
ccttcatttg tggctctatca atcaagtggg ctacgaatga ctctaatttt aagattatft	1080
ttgtagttgt gtggtgtfff agtagttacc aatcttatac ttgaaagaaa atgaaagcaa	1140
tgattactca tactactcaa tgccaagatc ggaggctaaa tccaatgtat acaagtatag	1200
aaatttgtaa agagttaagc tctttctttg ttcatgtagc tttgaggctt tgtaaaaata	1260
tggacattga ttcggatata gaggtgagtt gtgcacaaga gatgaccata cttggtgtca	1320
aggtgtagca tttttttcag attattttata agaaaataat caggaaagga aaataagtag	1380
tattcatcct agatataaca tttgtcgaga aatctacgag ataaacattt tttcagacga	1440
gaacaattct tcaaattttc agatgcaagg gtacgcattt agcattgcgc tgatattaga	1500
gctagtctcc tattgcatgt ttgatttcat acatgtacca cccattcttg ttactgcagt	1560
gtgtgaaact tgttgaataa gaagttccgc aattatttca aattattgag agtcttctta	1620
cataatfttt acttatccaa aattcttaag aaccccacia taaattcagt gatacgcttt	1680
gaatggctca ccagttactg gactgccaca attcgcagca ttggagactt ggccaactca	1740
accagagaag ggaccacgct gaacgatcta cctccctccc agtgagtgag tgagtcttcg	1800
ggtgcagtat tgtccaagtc ctggaatgtc gatccagccg caggaccagg aagatcgggc	1860
cggttacagt aaagttgcca taacaatccg gcaacgaacc acagatccgg gacgatctag	1920
cggaagttg aagttcaagg ctcggggcac atctccctgg tagaattaga atccatagcc	1980
agaattctat ctcgaaacct tgtttcgcca gcgttatgat tataatcaag cgtccccgtt	2040

aatctgattc ctgtgaaagt tagttagtaa cttcataccc cagcattatg attataatca	2100
agtgtctcag ttagtctgat tcctgtgaat gttagttagt aagttcaggc cttctcgtaa	2160
tagcttcttg cgtataatct gaactgttga taatgggttaa actcttgaat tacgacatat	2220
cagtcocggg agattaatct gcttcgcta agctcgagga tgcacagcag taattttggg	2280
tcgtttggga tttgataaaa cggacgggaa tatgcgtcgc gagttccgag taggagttag	2340
gaggaatgca aaccagcgga ccacgtaaag agggccacga cagtcagca gccagctgt	2400
gagacacaag ggggacgaaa gggacggccc agggcgacca cctgatgtca gggggagctg	2460
gtgcgagcgg cgacggacat ggatcggcgt ttggttgccg tccagaagcg ggcgaggagg	2520
gatccgcagt agtgacacag tgggggcaga attgggagaa gatcgtgggg gtaattgaga	2580
ggggagattc gggttggggc cgagacaggt aaggaacacc gatgatgctg aggaaaatat	2640
gaggaattcg tgagaatgcg acagggcgag agcactgtgg ggcagaatgg aaggggggccc	2700
agcgatattc gagcaataaa ataagagcgg gggacattcg aaaagaggcc ccatataaag	2760
ccgatcttcc attctgtttt cacagagctc ttcgtcgaac agagcctctc aaactcgctt	2820
tgtgctccca gtgcttctgt ctcgatctg ctctgctcgg cttcgcgctt gttgttcttg	2880
tgaccatcac cgccttcagg acgctcacgc ccaacgcaag aatttcgagt cgaagtaagc	2940
gagcagctca atcgcttcgt taacgcgttt gcggagatct tcgaggtttc gcgttcgaag	3000
ttcttcggac acctccttcg ttaac	3025

<210> 23  
 <211> 909  
 <212> DNA  
 <213> *Marchantia polymorpha*

<400> 23	
aagcttagca agcagctctc gcagcggatc tgctcttctg ctgctccctc tgcttcctcg	60
tgctacacgg tcttcgtcct cgcttcctcc acgcttcctc gcgctctctc caggtactcg	120
tcgcctcgcg ctctttcttc ttcctagttc gtccgttcct cgtaccggga tagggcggtc	180
gcgggtctcg tgagggtttt ttcgagcaag gtgcgtgagc aagttcatat cgggtgggcaa	240
tgcattggggc gaacctggtc gggccctttt ccgaggccgc cggagagcct agtctccaag	300
ctgtagtata ggtgttctcg aagatcggtc ggtgtctgca tctctccatc tcgattcgtt	360
tcgtctgagc tgatccgccg gtcgattttg acgatgtcgt gtcctcacct acgcaagttt	420
ggttccgagg attagttttg aagatgctgt caatgggaag tttagctctt ggttcgtgat	480
tagtttgagc acggtcacat gaatcgtagg gaccaggtg tcgggcggaa tcttcagcag	540
tcatttcggt ttcgtaacg ctggatttaa gctgaaaacg ttcacgatg gattgcggat	600
accatgacct aatggatcgt ccagcttatt cttctggaag tatagacgtg tgatggctgt	660

ggcctgtggt agggttggac acgcccgcag tggctctctcc gaatttgaat gtcgcaatgg	720
tcgatgtgct ctgccgattt ggggaatcga agtggcaaac cggtcgttcg gactgtcgag	780
tgtatgcctg ctgcttgtgc gatgtagtgt ggatttttcc tccgatgttt tccaaacgtg	840
gtcgggattg cagttcttca atctaccagc ggagctaatt tcgtctttgg cttgcagtct	900
atcgtcgat	909

<210> 24  
 <211> 2146  
 <212> DNA  
 <213> *Physcomitrella patens*

<400> 24	
atacaagagt tataaatcat atacaatgat tactttcata taattgttga atattattgt	60
tacaacctaa gtaacaataa cattcaatta aacattcatt gtggttttca agcatattaa	120
tcattctttc ttctctaccc tatagtgatg ggaaattatc ccaaactcaa tgtcatactc	180
caggcaattc agaaatatag tgagatgaat accaggaata ttatttcaca tcgaccctta	240
tcgccgggca atgccactcc caccgcggaa tgagaaactc cttgaaaaaa caagtccctt	300
cccagctgcc cgaaatcggc cgcctggtca gcacggcacg acaactgcca cgtgcaatcc	360
tgacgtggcc tctacgtccg gaaggcggcg ccgttagcga tgtcctccta tgcaagttcc	420
tcttgtggcg gggcagtggt cccgccaaact tcacgcgcac cctccacccc aacaagtggc	480
ccaaattact caggggcagc ccagcttcga aattttaagc ggtgaccgcc cttctctatc	540
gtcacgcgtt acttcttttt cactcaatcg agtctgttta ttattggccg ctaggaaatt	600
gcagcttcca actccgcac cccgcgtgca gtacagtggg gatcttcaag agtgtcctca	660
ccaggaattt gcaacttgct ccttgcaatt tgtaataaat ggacagagaa gcctagattc	720
cgcattccaca gtgatgggtc acgtatcaat aagcgaagct gcgttggtcaa ctatggcaat	780
tggtttggtg tcttcgttcc tgtcaagttt gaaaagaaga gggagatctg atttcttaat	840
aagtgtcgac ttgtctgggt agtggattgc gtggggcggtg tcgtagtgcg acgcgatcgc	900
atcaaattca tcgcctcaaa atttgtcacg ttgttgggtc aattgcaacg aactgcgatt	960
gaaggattct tctcgggtggc cttcaaattt gcttttagtat gacagaagtt ttgcagctgt	1020
actcggcggt tggaaggagt ggaagtgagg tggatcacca cgcaccggag ttggtgaatt	1080
gtttactgca gaaaaaatg gctttgatca catcagaatg attgatgttt cagcttgaat	1140
ttcacctcaa gatgtgttct catcatgaaa tttttattgg gccaggatgt actttcattg	1200
ttttgaaaga atattttaag acgcttgtgt tttaaacct ttcggaagat gcgtccttga	1260
ttgaaagtgg ttaatgtttt gtacatcatt actggatatg aaaataccaa taaaatgaaa	1320
tacaataaaa tatttttttg aaatgaaaat tggtttaaat aagcatgtaa ataatagacg	1380

gtggagtaaa gaaaaggtaa taaaaaaaaa agtatgaatt ctattactct tcaatataaa	1440
agtaagaggt gtccgtttgc aagcaataaa aattcagtaa ttgctagata aattcaaaag	1500
ccaaccaata cacaccattg ttttgctgca aagctagggt ttctaaggcc acaattcaat	1560
gactagtgc ttacatatta cttccaaacc gaagcaaagc aagggtactc caccgattgta	1620
tatatactca cttgttttatt tttaaaccat ctgaaatcac aaaaaaatgt tgtgaccctg	1680
cttcattatg ataattaagt gacgttttaa totcattaaa tttaatgcca ccgtaggtta	1740
tggacggaaa tggatggatg taaatggaaa gatcggcggc aaaaagacca aattccatac	1800
tactgcccga gtccgataaa gacggaaaca atgcgataaa agtaaaagtg agcagaagaa	1860
agtgcacggt cgaaggcggc gtttgtttac atttacttca ccaaaaccga gcaggatatac	1920
gggcacacgg tcaggaagaa attgttcatg acggtcagaa cattctggat ggttggcgtg	1980
cttgctataa gaacactgct cctccgatct aaacctcgga ttgtgcgctt ctagatactg	2040
aatttgtttc gaccctgcct tgttgagtgg ccgtagaggc tcgacagtta ggatcagtgt	2100
gccgttgaat ttagtgattg tgtagcgacc agtacgtcct gtaagg	2146

<210> 25  
 <211> 524  
 <212> DNA  
 <213> *Funaria hygrometrica*

<400> 25	
gaattcattt ccattaacga gaatatgaca gtgggaagag cttccacgtc atccaaaactc	60
aaagtatccg acgtgggtcaa tccaagtgcc agtgccacct cagctccttc accagtccat	120
ctcgcggata aggggtgacag caaggcgcggt tattactgga taagagaagc ggccaaggcg	180
gcagccactg tgggtccactt tgetgogtca ctacctactg cgattgtaat gacgagcggc	240
agcgctcgtgt gacaggcttg aaccgaccgc tgcttcagcc gcaggcagac tagaaaagtt	300
tactcgtgt cccactcgtt ttctgggtgt gcatccgaag ttctctggatg gttgcccgtc	360
gttcaataaa ttgtcgcgcg tcgagctagc ggacactttt gtcacogttc ttctctgttt	420
attctggacc agagggtgctg ttagctttgt tgtgtgtgag tccttgggga aatccctgcg	480
cgtcacgaga gtttattgca ggggaagtgat aaagcgttgt gaag	524

<210> 26  
 <211> 2088  
 <212> DNA  
 <213> *Physcomitrella patens*

<400> 26	
atgcatgtaa gataattcca attagaatct ataaatttct tattataatt ttttaaaaac	60
aaagtaccaa aatattatta ttttaatatc ctctaagtta aatccatata ttaagtagaa	120

acaattattc taataaataa tgataaaaat tagacatctt gcaataaaaat ttcttttttaa	180
aaatagatagc ataacatgaa aaatatccca taaatagcta acaccatcaa aacatttgac	240
caaatatgca ctttttagatg tgtcaagaca aaaagaaata tttgcaagat tttggagtat	300
ctaaactaat gtttgctctc tttgcactat gagtaggatt tctttttattt tgttttagtga	360
aaagatacat tgcaatttgt tttcataata aaaactatac taatgaaata gtgctaaaaa	420
ataacaagat taaaaaaaca taacccttct tacaacctta aatccttcta attagactac	480
ctcaaagttg tgccatttag cacaaaaacc attcttttaa atctacttaa cctccaatt	540
tccaatgagc ttcattgtgca tacacaagca tgctttcttt ctttctttct tgaagaaaac	600
ttatctgaac aaacgttaat actctacttg ttgatgaaag tggaactttg accacatata	660
ggcttggtga tgtactttgt atatctctc acagttagtc tgggtgcaatc caaccatgca	720
catagaatat gaatggggac atgcttccag ccactcgggt gtgcagaaaa cttgacaagc	780
gagattcaag caacggcgac tacgacgccg atcacgcaat acaaagcatt gttagtatgt	840
gataaaccag agaaagagat cgagtatgtg cacacaaaaa cacacagatc cacaggtatt	900
gtctacggcg ccaccaccat ccgtcaaagc taccatctcg tcgaggaaga atgggtatttc	960
taaaactagc aatacaaccg ctgatggaaa caaccgaaag ctatgtcatt ggagagggcg	1020
cacgagttca tggaatacac agtgagaaga gataaagaaa taaaataata taaaatacaa	1080
gtgtgcatca gcaagacatg gccgaaatct aacaactgtc tgcacatgct gtggtgggtt	1140
gtatccacgc gctggaggaa gtaactttcc tacatgcaca gaaaaacatt ttcagattag	1200
aaagctcttc tgttctagct aatctctagt accaagctca gacgtgtagc cgacgaagcc	1260
aatagcagct gggatatgcta gtcactgatt ctgaagcggc cgggtgtgtcg attgcatgt	1320
atctcagttc ggcgaaggcc tgtgtctgga acatgggaag aggtctctct tgcactcgtc	1380
aatctctcac agcaactggg cagggttgta tccgaacgtg gaaaacgcag caaccgttgt	1440
tgaaccaaag gatggtattt ttctccgaga aaaacgccgt ggcttatctg gtgtagacga	1500
tcctaatcc ggacatgacc gccgctgtgc aggtgttggg aaaccacaat gcgcaagaga	1560
tgcgagagat ggaggagtgc aagaagtacg actgcgaagc tacatgcttc atcgagcaat	1620
gaagtctggg ttttctccaa cttccgcatg cacacacttt tctcgacgac atccgtttca	1680
aggtacgcat cgggaaactg acgattctct gcaactggtg tcagactctc cggagaggcg	1740
gtgtcatgtt ctgagctctt tttcgataag gtgctgttga agtccagaat aatggggtct	1800
ggattatcct ctggacggct ccgcttctgg tcgaaaaaat ttcattccaa aaaaggactt	1860
atctgttgac tgaaaatgtt taattgtggt gaggattgca tgcagcgacg tcgtaaagat	1920
agggtgacaa ggagcgttcc agagctcagc tcggggcatg ccccggaact ccttagcata	1980
taaacatacc gggtggaatt tgtaccacc aggtcttgct cgggtgtccc tgtgccaag	2040

ctgttggctg cattgccctt gcgattcgag tgtggagaga ttttagca

2088

<210> 27

<211> 500

<212> DNA

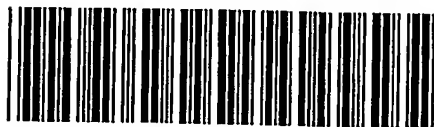
<213> *Physcomitrella patens*

<400> 27

ggaacgaatt tgtcgagctc tctggttctg ggtcgggtag cagtagcttt gatggtgagg	60
cactgacagt cagtcgctca cacggcaaag tagcctggat gtgcttcgca acgaactctt	120
gaatttgagt atgtgagttc actttgaaca tcccagaagc aaaagaatgg gttttttcat	180
gtttgaatth tttttgtat agttgtgttg agccgcgatt tctatctgtc acttggcttg	240
atattctgag tttctccgat acgaatagcg aagtccactt gaacatctgt aacggcagca	300
attgcgtcag gtcaatcctc tcagattctt tcggtgcttt tgtcgtaaac tagcttgatt	360
gttgtccatt aagcttggtt gcttttcgtg agaaagcatg aaacttctat gacgaaaccc	420
ggttgattgt aatgtaacta gtttgattgt agtttgaatt tggttaattgc gttgtatgat	480
acataatgaa agtttcatga	500



**PCT/EP2004/008580**



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☒ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**